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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 November 2023 (30.11.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/230483 A2

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/11* (2006.01)

(21) International Application Number:
PCT/US2023/067370

(22) International Filing Date:
23 May 2023 (23.05.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/344,896 23 May 2022 (23.05.2022) US
63/349,313 06 June 2022 (06.06.2022) US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(72) Inventors: **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **ALTAE-TRAN, Han**; c/o 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **KANNAN, Soumya**; c/o 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(74) Agent: **HARDING, Drew P.** et al.; Johnson, Marcou, Isaacs & Nix, LLC, Customer No. 95280, PO Box 691, Hoschton, Georgia 30548 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, 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).

(54) Title: ENGINEERED CHIMERIC ISCB POLYPEPTIDES AND USES THEREOF

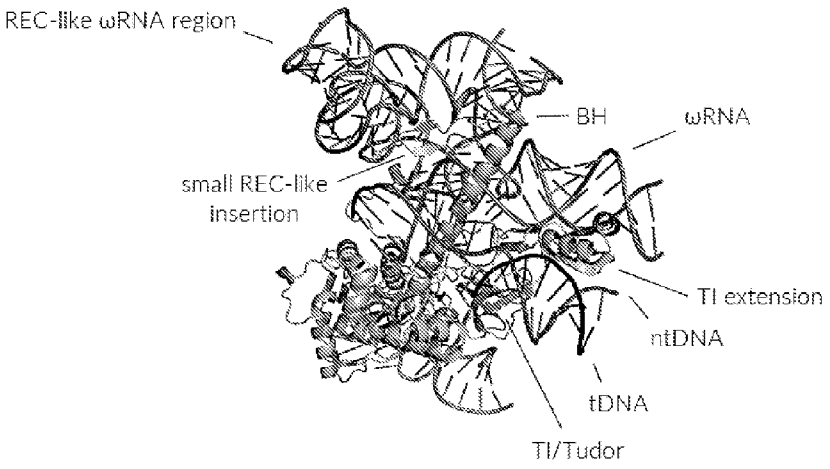


FIG. 1B

(57) Abstract: Chimeric, engineered DNA-targeting IscB systems, methods and compositions including novel chimeric IscB polypeptides and reprogrammable targeting nucleic acid components and methods and application of use are provided.

[Continued on next page]

WO 2023/230483 A2

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))*

ENGINEERED CHIMERIC ISCB POLYPEPTIDES AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/344,896, filed May 23, 2022 and U.S. Provisional Application No. 63/349,313, filed June 6, 2022. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant Nos. HL141201 and HG009761 awarded by The National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] The contents of the electronic sequence listing ("BROD-5590WP_ST26.xml"; Size is 34,258,745 bytes and it was created on May 23. 2023) is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0004] The subject matter disclosed herein is generally directed to systems, methods and compositions used for targeted gene modification and nucleic acid editing utilizing systems comprising engineered IscB compositions. In particular, compositions comprising chimeric IscB polypeptides with increased specificity or activity relative to unmodified IscB polypeptides are provided.

BACKGROUND

[0005] IscB compositions comprising IscB polypeptides and an ωRNA molecule that can be engineered to direct sequence specific binding of the IscB polypeptide to a target polynucleotide. Further improvement of the systems, including increased specificity and activity would be additional desirable tools in genome engineering and biotechnology that would further advance the art.

[0006] Citation or identification of any document in this application is not an admission that such a document is available as prior art to the present invention.

SUMMARY

[0007] In one aspect, as described herein, an engineered IscB composition comprising: a) an IscB polypeptide comprising one or more insertions of a heterologous polypeptide, and optionally one or more modified amino acids, that increases specificity or activity of the chimeric IscB polypeptide relative to wildtype; and b) an ωRNA molecule comprising a scaffold and a reprogrammable spacer sequence, the ωRNA molecule capable of forming a complex with the IscB polypeptide and directing sequence-specific binding of the IscB polypeptide to a target polynucleotide.

[0008] In an example embodiment, the insertion is a Rec domain, or functional fragment thereof. In an example embodiment, the Rec domain is from a Type II Cas polypeptide. In an example embodiment, the Rec domain is a Rec domain from a Type II-D Cas polypeptide. In an example embodiment, the Rec domain is a Rec domain from a Cas9. In an example embodiment, the Cas9 is derived from Francisella novicida (FnoCas9), Neisseria meningitidis (NmeCas9), Staphylococcus aureus (SaCas9), Streptococcus pyogenes (SpCas9), Streptococcus thermophilus (StCas9), Acidothermus cellulolyticus (AceCas9), Campylobacter jejuni (CjeCas9) or a combination thereof. In an example embodiment, the Rec domain is inserted between amino acids 153-160 of Rd8_117 polypeptide from Table 1, or an analogous position of another IscB polypeptide.

[0009] In an example embodiment, the ωRNA comprises a deletion that reduces steric interference with the inserted Rec domain. In an example embodiment, the deletion is in the PK-loop of the ωRNA. In an example embodiment, the deletion comprises 1 to 30 nucleotides of SEQ ID NO: (reference ωRNA) or an analogous position in another ωRNA.

[0010] In an example embodiment, the insertion is a protein capable of binding to RNA, DNA, or both. In an example embodiment, the insertion is a nuclease, or a functional fragment thereof. In an example embodiment, the insertion is an endonuclease, exonuclease, or functional fragment thereof. In an example embodiment, the endonuclease is a Ribonuclease (RNase), deoxyribonuclease (DNase), or fragment thereof. In an example embodiment, the insertion is hybrid binding domain (HBD). In an example embodiment, the insertion is a RuvC

domain or portion thereof. In an example embodiment, the RuvC domain comprises a Cas9 RuvC domain/region, subdomain/subregion, or portion thereof.

[0011] In an example embodiment, the insertion is a TAM interacting (TI) or PAM interacting (PI) domain, or functional fragment thereof. In an example embodiment, the domain is an NGG PI or TI domain, or functional fragment thereof. In an example embodiment, TAM determining region comprises one or more amino acid substitutions. In an example embodiment, the insertion is in the WED/adaptor stabilizer region, the Tudor domain, Tudor Lance domain, or a combination thereof. In an example embodiment, the insertion preserves RNA interaction with the TAM determining region. In an example embodiment, the TI domain, PI domain, or functional fragment thereof is inserted between amino acids 365-499 of Rd8_117 from Table 1, or an analogous position of another IscB polypeptide. In an example embodiment, the TI domain, PI domain, or functional fragment thereof is from a Type II Cas polypeptide. In an example embodiment, the TI domain or functional fragment thereof is from an IscB polypeptide.

[0012] In an example embodiment, the insertion replaces amino acids 365-499, 369-499, 462-486 (Lance), 450-486 (Tudor), or 376-383, 436-448, and 462-486 (Lance_TIL) of the TI domain of IscB polypeptide Rd8_117 from Table 1, or an analogous position of another IscB polypeptide. In an example embodiment, the TAM of the wild-type IscB polypeptide is retained or wherein the TAM is modified. In an example embodiment, the insertion comprises one or more amino acid positions from 380-735 from SEQ ID NO: 2365 one or more amino acid positions 556-609 from cA2 ProCas9-2, one or more amino acid positions 356-420 from IscB_Rd8_149, one or more amino acid positions 386-464 from IscB_Rd4_7, one or more amino acid positions 856-924 from Cas9_971, one or more amino acid positions 569-751 from Cas9_1079_3, one or more amino acid positions from ChlorIscB, one or more amino acid positions 488-512 from SEQ ID NO: 2367, one or more amino acid positions 739-765 from SEQ ID NO: 2365, one or more amino acid positions 407-431 from IscB_Rd8_127, one or more amino acid positions 404-430 from CRISPR IscB 00644, one or more amino acid positions 376-482 from IscB_large_28, one or more amino acid positions 356-488 from IscB_Rd8_149, one or more amino acid positions 376_482 from IscB_Rd8_75, one or more amino acid positions 374-495 from IscB_Rd8_151, one or more amino acid positions 376-477 from IscB_Rd8_23, one or more amino acid positions 376-477 from IscB_Rd8_24, one or more amino acid positions 376-486 from IscB_Rd8_118, one or more amino acid positions 374-495

from IscB_Rd8_151, and/or one or more amino acid positions 376-486 amino acids from IscB_Rd8_118.

[0013] In an example embodiment, one or more nucleotides in a pseudoknot nexus of the ωRNA is modified and/or inserted. In an example embodiment, one or more nucleotides in a nexus stem of the ωRNA that base pair to the one or more nucleotides in the pseudoknot nexus is modified and/or inserted. In an example embodiment, a base pair comprising a nucleotide in the pseudoknot nexus and a nucleotide in the nexus stem is substituted with a complementary base pair. In an example embodiment, one or more nucleotides in a pseudoknot region of the ωRNA are modified and/or inserted and wherein the nexus stem retains base pairing. In an example embodiment, one or more nucleotides in a pseudoknot region of the ωRNA are modified and/or inserted and wherein the pseudoknot retains its structure relative to a wild-type IscB. In an example embodiment, the pseudoknot region of the ωRNA retains its structure by no less than 50%, no less than 55%, no less than 60%, no less than 65%, no less than 70%, no less than 75%, no less than 80%, no less than 85%, no less than 90%, no less than 95% relative to a wild-type IscB. In an example embodiment, the pseudoknot comprises a peptide nucleic acid (PNA).

[0014] In an example embodiment, the 3'-end of the ωRNA is truncated. In an example embodiment, the 3'-end of the ωRNA is truncated by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides. In an example embodiment, a RNA supplied in trans is bound to a 3'-end of the ωRNA. In an example embodiment, one or more amino acids in contact with a nucleotide are substituted or removed. In an example embodiment, one or more amino acids in contact with the ωRNA are substituted or removed. In an example embodiment, one or more amino acids on the surface of the engineered IscB are substituted or removed.

[0015] In an example embodiment, the one or more amino acids on the surface of the engineered IscB are substituted to have a different charge. In an example embodiment, the one or more amino acids on the surface of the engineered IscB are removed or substituted to increase or decrease the overall charge of the protein surface. In an example embodiment, the surface charge of the engineered IscB is more neutral relative to the wildtype.

[0016] In an example embodiment, the IscB further comprises a nucleotide deaminase. In an example embodiment, the insertion is a nucleotide deaminase. In an example embodiment, the nucleotide deaminase is an adenosine deaminase or cytidine deaminase. In an example

embodiment, the nucleotide deaminase is inserted in the middle of the IscB or fused at the N terminus or C-terminus of the IscB polypeptide. In an example embodiment, the cytosine deaminase is apolipoprotein B mRNA-editing enzyme, catalytic polypeptide (APOBEC), an activation-induced deaminase (AID), a cytidine deaminase 1 (CDA1), or cytosine deaminase acting on RNA (CDAR). In an example embodiment, the adenosine deaminase is ADAR or TadA. In an example embodiment, the adenosine deaminase is an ADAR and the reprogrammable spacer sequence of the ωRNA molecule comprises one or more mismatches to the target polynucleotide. In an example embodiment, the C-terminus is engineered for base-editing. In an example embodiment, the insertion is a transposase. In an example embodiment, the transposase is a TnpA.

[0017] In an example embodiment, the engineered IscB further comprises a functional domain. In an example embodiment, the functional domain comprises a base editing system or fragment thereof. In an example embodiment, the functional domain comprises a prime editing system or fragment thereof. In an example embodiment, the functional domain comprises an epigenetic editing system or fragment thereof. In an example embodiment, the functional domain is inserted at a junction. In an aspect, described herein, is a method of modifying a target polynucleotide comprising contacting a cell with a composition of any of the preceding claims.

[0018] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0020] **FIG. 1A-1B** – (1A) Ribbon diagram of CjeCas9 domains, including REC helical bundle involved in RNA stem binding; (1B) Diagram of IscB domains, including REC helical bundle involved in RNA stem binding.

[0021] **FIG. 2A-2B** – (2A) Ribbon diagram showing early REC-domains have IscB-L-compatible join locations; (2B) sequence diagramming junctions for insertion of REC domains, with insertions within positions 153-160 on Rd8_117 IscB.

[0022] **FIG. 3** – Ribbon diagram showing insertion of example II-D Cas9 REC3 domain into IscB.

[0023] **FIG. 4A-4C-** (4A) Includes ribbon diagram showing example PK-loop of ωRNA clashes with many REC domains, designed insertion location shown; (4B) ribbon diagram depicting REC-like ωRNA region; (4C) highlighted REC-like ωRNA region in ribbon diagram with approach to remove clashing region and rejoin with linker.

[0024] **FIG. 5A-5B** – (5A) Ribbon diagram including small region of CjeCas9 REC fragment inserted into Rd8_117 shows classh-free folding; (5B) Alphafold2 models show clash-free folding for Rd8_117 IscB + CjeCas9 REC fragment.

[0025] **FIG. 6** – Alphafold2 model showing new RNA/DNA hybrid recognizing groove in example IscB polypeptide.

[0026] **FIG. 7** – Ribbon diagram of RNaseH insertion in same region of example IscB shows fit, but that may need some rearrangement; this region of IscB can also permit other domain insertions.

[0027] **FIG. 8** –FnCas9 guide RNA circularly reroutes ωRNAs.

[0028] **FIG. 9** – Example IscB ribbon diagram showing 2 TAM determining regions, with WED/adaptor stabilizer region and Rud domain region highlighted.

[0029] **FIG. 10** – Example IscB ribbon diagram with strong polar and hydrophobic interactions suggesting reduction in TAM length could present challenge.

[0030] **FIG. 11** – Ribbon diagram of IscB polypeptide indicated large structural diversity exists in both the WED/adaptor stabilizer domain and the tudor core+divergent extensions.

[0031] **FIG. 12** – Ribbon diagram for example IscB polypeptide Rd8_117 that has3 key contacts with TAM-proximal DNA.

[0032] **FIG. 13A-13F** – Example IscB polypeptide engineering approaches (13A) full TI domain exchange with an NGG TAM/PAM; (13B) RNA interaction-preserving insertions with large IscBs which can include insertion at one linker, two linkers, and/or the tudor lance region which can correspond to Rd8_117 amino acid positions 376-383, 446-448, and 462-486; (13C) exchange of Tudor Lance region with NGG TAM/PAMs; (13D) exchange of Tudor region with NGG TAM/PAMs; (13E) depicts more comprehensive insertions with domains of IscB

polypeptide Rd8_149; (13F) depicts more comprehensive insertions with domains of IscB polypeptide Diverse_7.

[0033] **FIG. 14** – Depicts IscB polypeptide Rd8_117 C-terminal helix interacting with nucleotide strand.

[0034] **FIG. 15** – Depiction of pseudoknot (PK) region of ωRNA as a transRNA off-switch.

[0035] **FIG. 16** – Depiction of terminal hairpin region of ωRNA as a transRNA on-switch.

[0036] **FIG. 17** – Includes gel showing chimeric IscBs with Cas9 REC domains are functional in vitro.

[0037] **FIG. 18** – Includes gel of chimeric IscBs with Cas9 REC domains to extend enforced guide:target duplex are functional in vitro; REC domain insertions in Rd8_117 IscB.

[0038] **FIG. 19** – Charts chimeric IscBs with Cas9 REC domains mediate indel formation in human cells (Rd8_117) % indels for protein/REC insertion for guide 1 and guide 2.

[0039] **FIG. 20A-20B** – (20A) REC insertions destabilize TAM-distal guide mismatches for example guide mismatches; (20B) results of guides with varying mismatches.

[0040] **FIG. 21** – 1261 REC graft in example IscB polypeptide accesses more target sites and improves indel activity in general at varying loci.

[0041] **FIG. 22A-22F** – (22A) Weblogo shows no changes in TAM of example Rd8_117 IscB polypeptide with insertion of full TI domain; (22B) weblogo shows no changes in TAM of example Rd8_117 IscB polypeptide with partial domain insertion; (22C) weblogo shows no changes in TAM of example Rd8_117 IscB polypeptide with Tudor domain insertion; (22D) weblogo shows no changes in TAM of example Rd8_117 IscB polypeptide with Lance domain insertion; (22E) weblogo shows some Lance TIL domain insertions retain same TAM in example Rd8_117 IscB polypeptide; (22F) weblogo shows insertion of Lance_TIL domain can lead to alterations in TAM of example Rd8_117 IscB polypeptide.

[0042] **FIG. 23**– Weblogo of altered TAMs in comparison to TAM from wild-type Rd_117 IscB polypeptide.

[0043] **FIG. 24A-24D** – ωRNA engineering in example Rd8_117 IscB polypeptide. (24A) Rd8_117 can tolerate up to 21 nucleotides truncated from the 3' end of the ωRNA; (24B-C) Insertions in nexus pseudoknot are minimally functional; (24D) Insertions in nexus pseudoknot can be rescued by compensating insertions in nexus stem.

[0044] **FIG. 25** – Sequence map for Rd8_117_REC_SpCas9_RuvC_loop insertion.

[0045] **FIG. 26** – Sequence map for Rd8_66_SpyCas9_Hyb_Stab_ins1.

[0046] **FIG. 27** – Sequence map for Rd8_117_2089_REC_ins1.

[0047] **FIG. 28** – Sequence map for Rd8_117_Ace_REC_ins1.

[0048] **FIG. 29** – Sequence map for Rd8_117_Cas9_665_REC_ins1.

[0049] **FIG. 30** – Sequence map for Rd8_117_Cas9_971_1_REC_ins1.

[0050] **FIG. 31** – Sequence map for Rd8_117_Cas9_971_2_REC_ins1.

[0051] **FIG. 32** – Sequence map for Rd8_117_Cas9_1079_1_REC_ins1.

[0052] **FIG. 33** – Sequence map for Rd8_117_Cas9_1079_2_REC_ins1.

[0053] **FIG. 34** – Sequence map for Rd8_117_Cas9_1079_3_REC_ins1.

[0054] **FIG. 35** – Sequence map for Rd8_117_Cas9_1079_4_REC_ins1.prot.

[0055] **FIG. 36** – Sequence map for Rd8_117_Cas9_1261_REC_ins1.

[0056] **FIG. 37** – Sequence map for Rd8_117_Cdi_REC_ins1.

[0057] **FIG. 38** – Sequence map for Rd8_117_Cje_REC_ins1.

[0058] **FIG. 39** – Sequence map for Rd8_117_Fno_REC_ins1.prot.

[0059] **FIG. 40** – Sequence map for Rd8_117_Fno_REC_ins2.

[0060] **FIG. 41** – Sequence map for Rd8_117_Fno_REC_ins3.

[0061] **FIG. 42** – Sequence map for Rd8_117_Fno_REC_ins3.

[0062] **FIG. 43**– Sequence map for Rd8_117_Nme1_HNH_swap1.

[0063] **FIG. 44** – Sequence map for Rd8_117_Nme1_REC_ins1.

[0064] **FIG. 45** – Sequence map for Rd8_117_Nme1_RuvCIII_ins1.

[0065] **FIG. 46** – Sequence map for Rd8_117_Rd8_66_REC_ins1.

[0066] **FIG. 47** – Sequence map for Rd8_117_Sau_REC_ins1.

[0067] **FIG. 48** – Sequence map for Rd8_117_Spy_REC_ins1.

[0068] **FIG. 49** – Sequence map for Rd8_117_SpyCas9_Hyb_Stab_ins1.

[0069] **FIG. 50** – Sequence map for Rd8_117_Sth_HNH_swap1.

[0070] **FIG. 51** – Sequence map for Rd8_117_Sth_REC_ins1.

[0071] **FIG. 52** – Sequence map for full TI exchange Rd8_117_REC_665_TI.

[0072] **FIG. 53** – Sequence map for full TI exchange Rd8_117_REC_1079_1_TI.

[0073] **FIG. 54** – Sequence map for full TI exchange Rd8_117_REC_1079_2_TI.

[0074] **FIG. 55** – Sequence map for full TI exchange Rd8_117_REC_1079_3_TI.

[0075] **FIG. 56** – Sequence map for full TI exchange Rd8_117_REC_2089_TI.

[0076] **FIG. 57** – Sequence map for full TI exchange Rd8_117_REC_cA2_TI.

[0077] **FIG. 58** – Sequence map for full TI exchange Rd8_117_REC_Diverse_7_TI.

[0078] **FIG. 59** – Sequence map for full TI exchange Rd8_117_REC_Rd8_149_TI.

[0079] **FIG. 60** – Sequence map for partial TI insertion Rd8_117_REC_644_2_tudor_ext.

[0080] **FIG. 61** – Sequence map for partial TI insertion Rd8_117_REC_644_2_tudor_lance.

[0081] **FIG. 62** – Sequence map for partial TI insertion Rd8_117_REC_644_2_tudor.

[0082] **FIG. 63** – Sequence map for partial TI insertion Rd8_117_REC_2089_tudor_lance.

[0083] **FIG. 64** – Sequence map for partial TI insertion Rd8_117_REC_2089_tudor.

[0084] **FIG. 65** – Sequence map for partial TI insertion Rd8_117_REC_cA2_tudor_lance.

[0085] **FIG. 66** – Sequence map for partial TI insertion Rd8_117_REC_cA2_tudor_swap.

[0086] **FIG. 67** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_665_tudor_lance.

[0087] **FIG. 68** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_665_tudor.

[0088] **FIG. 69** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_971_tudor.

[0089] **FIG. 70** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1079_1_tudor_lance.

[0090] **FIG. 71** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1079_1_tudor.

[0091] **FIG. 72** – Sequence map for partial TI insertion. Rd8_117_REC_Cas9_1079_2_tudor_lance.

[0092] **FIG. 73** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1079_2_tudor.

[0093] **FIG. 74** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1079_3_tudor_lance.

[0094] **FIG. 75** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1079_3_tudor.

[0095] **FIG. 76** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1261_tudor_lance.

[0096] **FIG. 77** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1261_tudor_lance.

[0097] **FIG. 78** – Sequence map for partial TI insertion Rd8_117_REC_Cas9_1261_tudor.

[0098] **FIG. 79** – Sequence map for partial TI insertion Rd8_117_REC_ChlorIscB_tudor_lance.

[0099] **FIG. 80** – Sequence map for partial TI insertion Rd8_117_REC_ChlorIscB_tudor.

[0100] **FIG. 81** – Sequence map for partial TI insertion Rd8_117_REC_Diverse_7_NTS.

[0101] **FIG. 82** – Sequence map for partial TI insertion Rd8_117_REC_Diverse_7_REC.

[0102] **FIG. 83** – Sequence map for partial TI insertion Rd8_117_REC_Diverse_7_TI_L_tudor_lance.

[0103] **FIG. 84** – Sequence map for partial TI insertion Rd8_117_REC_Diverse_7_TI_L.

[0104] **FIG. 85** – Sequence map for partial TI insertion Rd8_117_REC_Diverse_7_TI_L1.

[0105] **FIG. 86** – Sequence map for partial TI insertion Rd8_117_REC_Diverse_7_TI_L2.

[0106] **FIG. 87** – Sequence map for partial TI insertion Rd8_117_REC_Diverse_7_TI_restore_WED.

[0107] **FIG. 88** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_127_tudor_lance.

[0108] **FIG. 89** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_149_NTS_1.

[0109] **FIG. 90** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_149_NTS_2.

[0110] **FIG. 91** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_149_TI_L.

[0111] **FIG. 92** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_149_TI_restore_WED.

[0112] **FIG. 93** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_149_tudor_nontudor.

[0113] **FIG. 94** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_149_WED_ins.

[0114] **FIG. 95** – Sequence map for partial TI insertion Rd8_117_REC_Rd8_149_WED_swap.

[0115] **FIG. 96A-96E** – Biochemical characterization of IscB polypeptides: 96A) gel of biochemical measurements with varying RNA:protein molar ratio from 4 to 0.25 for Rd8_66, Rd8_117, and Rd8_117 + 1079_1 REC insertions; 96B) activity of IscBs Rd8_66, Rd8_117, and Rd8_117 _1079_1 REC insertion by temperature varying between 22C and 67C; 96C) RNA modification can modulate activity at different temperatures; 96D) cleavage kinetics measured for Rd8_66, Rd8_117, and Rd8_117 + 1079_1 REC insertions at times varying

between 0 and 120 minutes; 96E) characterization of divalent metal ion on cleavage activity of Rd8_66, Rd8_117, and Rd8_117 + 1079_1 REC insertions.

[0116] **FIG. 97** – Rd8_117 with REC insert can be functionally delivered by eVLPs to HEK293 cells.

[0117] **FIG. 98A-98C** – 98A) Example target base numbering for Adenine base editing with example ωRNA guide sequence; 98B) example Rd8_117 IscB polypeptide is compatible with ABE8 adenine base editing, as measured by editing at several target loci; 98C) example chimeric Rd8_117 IscB polypeptide with 1079_1 REC insert shows improved ABE8 adenine base editing activity as measured by editing at several target loci.

[0118] **FIG. 99A-99B** – 99A) Rd8_66 and Rd8_117 demonstrate detectable binding of DNA via PAM -SCANR; 99B) Rd8_66 indel formation based on variations in structure of the ωRNA at VEGFA1 loci.

[0119] **FIG. 100** – Indel rates measurement across guides with a 3' 20bp truncation with Rd8_66.

[0120] **FIG. 101** - Schematic of chimeric ωRNA engineered for trans on-switch and trans off-switch.

[0121] **FIG. 102A-102D** – 102A) Rd8_117 cleavage activity measured with ωRNA 3' truncations; 102B) schematic of ωRNA truncations for 102A; 102C) Rd8_117 cleavage activity measured with ωRNA 5' truncations; 102D) schematic of ωRNA truncations for 102C.

[0122] **FIG. 103A-103B** – 103A) Results of Rd8_117 nexus loop reprogramming; 103B) nexus loop sequence modifications.

[0123] **FIG. 104A-104B** – 104A) Depiction of nexus pseudoknot insertion variants relative to WT ωRNA; 104B) cleavage activity of ωRNA variants.

[0124] **FIG. 105** – Depictions of ωRNA folding with 21 bp 3' truncation and relative to full ωRNA with Gibbs free energy measurements.

[0125] **FIG. 106A-106D** – 106A) Addition of the last 35 base paris of Rd8_117 ωRNA can restore activity to the Rd8_117 35 bp 3' truncated ωRNA; 106B-C) 35 bp transRNA is active with multiple guides using example IscB Rd8_117 protein ((106B) and Rd8_117 +1079 REC insertion (106C) m = 14 bp that span distance from -35 from 3' to -21 from 3' end of ωRNA; 106D) .Rd8_117 guide scaffold structure with truncations indicated.

[0126] **FIG. 107A-107B** - Trans-on RNA restores activity in mammalian cells, WT is the unmodified Rd8_117 IscB; 1079 is Rd8_117 IscB with Cas 9 1079_1 REC insertion; full

transRNA = last 35 bp of Rd8_117 ωRNA scaffold; +G is addition of G after U6 and before the transRNA sequence; mini transRNA = 14 bp that space distance from 35 from 3' to -21 from 3'. Condit for 107A) DYNCH1 and 107B) HPRT1. Conditions not containing 'notation 'full' contain a 3' end 35 bp truncation of the Rd8_117 ωRNA.

[0127] **FIG. 108** – Multiple allosterically distinct mechanism are improved from natural mutations in IscB.

[0128] **FIG. 109** – Hydrophobic repacking of RuvC improves IscB activity by 2x.

[0129] **FIG. 110** – Extending the DNA/RNA duplex channel via amino acid modifications in IscB improves activity 2x.

[0130] **FIG. 111** – Creating a new non-targeting strand channel by amino acid mutations of IscB polypeptide improves 1.7-1.9x.

[0131] **FIG. 112** – Altering unwinding activity via pinpoint mutations improves IscB activity 2x.

[0132] **FIG. 113** –Improvement from enhancing existing DNA/RNA channel with mutations to IscB polypeptide.

[0133] **FIG. 114** – Non-specific DNA binding mutations improve IscB activity by 2x.

[0134] **FIG. 115** – New IscB mutants can be made by combining mutations, including those identifies in FIGS. 108-114 affecting different mechanism.

[0135] **FIG. 116A-116B** – Alignment of REC-flanking regions in IscBs and Cas9s. Alignment of (116A) N-terminal and (116B) C-terminal flanking regions of REC domains and REC-like inserts in IscBs and Cas9s. Residues are colored by BLOSUM45 similarity and residues with notable conservation in II-D Cas9 REC domains are marked with red triangles. Gray shaded regions denote REC domain regions.

[0136] **FIG. 117A-117F** – Structure and evolution-guided engineering of OrufIscB. (117A) Schematic of IVTT REC insertion screen. Each chimeric OrufIscB was incubated with a library of 100 guides targeting human genomic sites and a synthetic target library containing each possible single mismatch in the target and TAM and progressive mismatches on the TAM-distal end of the target in addition to the perfectly matched target. Cleaved targets were deep sequenced and cleavage was quantified based on read count. (117B) Median read count normalized to perfectly matched targets for TAM-distal mismatch targets. Only REC insertions with decreased cleavage at 5, 6 and 7 TAM-distal mismatch targets are displayed. (117C) Genome editing by wild-type OrufIscB and with the addition of the NbaCas9-1 REC domain.

Bars represent mean of n=4 replicates with error bars +/- standard deviation. (117D) Genome editing by aa mutants on top of enOrufIscB-v1 with two guides. Points represent mean of n=4 replicates at each guide with error bars +/- standard deviation. (117E) AlphaFold2 structure of wild-type OrufIscB overlayed on ωRNA and target DNA from the cryo-EM structure of OgeuIscB (PDB: 7UTN)262. Residues with selected mutations are highlighted in red. (117F) Comparison of genome editing by WT OrufIscB, enOrufIscB-v1, and all possible single, double and triple combinations of E137K, E409R and I533K variants with 12 guides. Bars represent mean of n=4 replicates with error bars +/- standard deviation.

[0137] **FIG. 118** – Comparison of genome editing by top candidate REC insert-containing OrufIscB variants. Genome editing by 54 REC insert- containing OrufIscB variants across 6 target sites in the human genome. Blue bar represents WT OrufIscB and red bar represents selected NbaCas9-1 REC domain-containing OrufIscB. Red line denotes indel efficiency of WT OrufIscB. Bars represent n=1 replicate.

[0138] **FIG. 119A-119B** – Guide length preference of OrufIscB. Genome editing at 4 sites in the human genome using ωRNA guides from 12 nts to 28nts with (A) OrufIscB and (B) enOrufIscB-v1. Bars represent mean of n=3 replicates with error bars +/- standard deviation.

[0139] **FIG. 120A-120C** – Base editing and transcriptional repression with enOrufIscB-v2. (120A) Schematic of rAAV genome design for enOrufIscB-v2 fusion platform packaging. (120B) Adenine base editing at 10 sites in the human genome with enOrufIscB-v2 using ABE8e in HEK293FT cells. (120C) Transcriptional repression of three genes in the human genome by OMEGAoff and CRISPRoff-v2.1 in HEK293FT cells. Fold expression is calculated relative to a non-targeting guide control.

[0140] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0141] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and

Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0142] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0143] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0144] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0145] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[0146] As used herein, a "biological sample" may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a "bodily fluid". The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female

ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0147] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[0148] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0149] The term "functional fragment thereof" should be interpreted as meaning any fragment having a desired function. For example, if a whole protein modulates enzymatic activity, then a functional fragment is a fragment capable of modulating enzymatic activity. In another example, if a whole protein breaks bonds in a molecule, then the functional fragment is a fragment capable of breaking bonds in a molecule. The exact quantitative function of the

functional fragment may be different from the function of the whole-size molecule. In some instances, the function of the functional fragment may be increased compared to the whole-size molecule. The use of fragments instead of whole-sized molecules can be advantageous because of the smaller size of the fragments.

[0150] Reference is made to International Patent Application PCT/US2021/056361 entitled Reprogrammable IscB Polypeptides and Uses Thereof, filed October 22, 2021, U.S. Provisional Application No. 63/105,191, filed October 23, 2020, entitled "Reprogrammable IscB Polypeptide Nucleases and Use Thereof", U.S. Provisional No. 63/105,177, filed October 23, 2020, entitled "Nucleic Acid-Guided Nucleases and Use Thereof," U.S. Provisional Application No. 63/156,857, filed March 4, 2021, entitled "Reprogrammable IscB Polypeptide Nucleases and Use Thereof", U.S. Provisional Application No. 63/195,659, filed June 1, 2021, entitled "Reprogrammable IscB Polypeptide Nucleases and Use Thereof", and U.S. Provisional Application No. 63/235,583, filed August 20, 2021, entitled "Reprogrammable IscB Polypeptide Nucleases and Use Thereof."

[0151] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

OVERVIEW

[0152] IscB systems are RNA-guided re-programmable nucleases. Altae-Tran and Kannan *et al.*, *Science* 374, 57-65 (2021). An IscB system comprises a IscB polypeptide and a nucleic acid component capable of forming a complex with the IscB polypeptide and directing the complex to a target polynucleotide. The IscB polypeptides are relatively small, typically ~400 amino acids, with a relatively large nucleic acid component, termed an ωRNA which comprises a guide sequence and a scaffold that interacts with the IscB polypeptide at several locations along the IscB polypeptide. Embodiments disclosed herein are directed to chimeric IscB systems that comprise one or more modifications that modify the IscB polypeptide, the ωRNA, or both. The modifications to the IscB comprise the incorporation of domains and other elements either from different IscB orthologs or from non-IscB proteins that increase the specificity or activity of the IscB polypeptide relative to wild type. Modifications to the ωRNA include additions, truncations, or insertions of heterologous polynucleotide sequences into the

scaffold portion of the ωRNA that reduce steric interference, enhance base pairing for increased specificity or activity, or provide elements that may be used to switch on or off IscB activity.

ISCB POLYPEPTIDES

[0153] Unmodified IscB polypeptides may comprise a N-terminal PLMP domain, a RuvC endonuclease, and a HNH domain. The RuvC domain may be a split RuvC domain comprising RuvC-I, RuvC-II, and RuvC-III subdomains. A bridge helix domain may be inserted between two of the RuvC domains. In one example embodiment, the bridge helix domain is inserted between the RuvC-I and RuvC-II subdomains. Unlike Cas9, IscB polypeptides do not contain a Rec domain. IscB proteins may also further comprise a conserved C-terminal domain. See, Schuler and Hu et al., "Structural basis for RNA-guided DNA cleavage by IscB-ωRNA and mechanistic comparison with Cas9," Science 10.1126/science.abq7220 (2022), incorporated herein by reference in its entirety.

[0154] In example embodiments, the unmodified IscB polypeptides are between 180 and 800 amino acids in size, between 200 and 790 amino acids in size, between 200 and 780 amino acids in size, between 200 and 770 amino acids in size, between 200 and 760 amino acids in size, between 200 and 750 amino acids in size, between 200 and 740 amino acids in size, between 200 and 730 amino acids in size, between 200 and 720 amino acids in size, between 200 and 720 amino acids in size, between 200 and 710 amino acids in size, between 200 and 700 amino acids in size, between 200 and 690 amino acids in size, between 200 and 680 amino acids in size, between 200 and 670 amino acids in size, between 200 and 660 amino acids in size, between 200 and 650 amino acids in size, between 200 and 640 amino acids in size, between 200 and 630 amino acids in size, between 200 and 620 amino acids in size, between 200 and 610 amino acids in size, between 200 and 600 amino acids in size, between 200 and 590 amino acids in size, between 200 and 580 amino acids in size, between 200 and 570 amino acids in size, between 200 and 560 amino acid, between 200 between 550 amino acids, between 200 and 540 amino acids, between 200 and 530 amino acids, between 200 and 520 amino acids, between 200 and 510 amino acids, between 200 and 500 amino acids, between 200 and 490 amino acids, between 200 and 480 amino acids, between 200 and 470 amino acids, between 200 and 460 amino acids, between 200 and 450 amino acids, between 200 and 440 amino acids, between 200 and 430 amino acids, between 200 and 420 amino acids, between 200 and 410 amino acids, between 200 and 400 amino acids, between 300 and 400 amino acids. between 300 and 500 amino acids, between 300 and 600 amino acids, between 400 and 500 amino acids,

or between 500-600 amino acids. In one example embodiment, the polypeptide may range in size from 400-500 amino acids, 400-490 amino acids, 400-480 amino acids, 400-470 amino acids, 400-460 amino acids, 400-450 amino acids, 400-440 amino acids, 400-430 amino acids. Size variation may be dependent, in part, on the particular domain architecture of the IscB or its homolog.

[0155] Table 1 – Example IscB system

SEQ ID 25276 (IscB RD8 117)	MRYVYVLDVDGKPLMPTCRFGKVRRMLKSGQAKAVDTLPFTIQLTY KPRTRILQPVTLGQDPGRTNIGMAAVRFDGKELGRFHCITRNKEIPKL MADRMAARKASRRGERLARKRLARKLHTTAKHLNGRILPGCSEPIAV KDIINTESRFNNRLRPEGWLTPTATQLLRTHINLFRKLSGILPVTDVAV ELNKFAFMQLDNPEMKKREIDFCHGPLCGTGGLEAAVKEQQDGKCL LCGKESIGHYHHIVPRSRRGSNTIANIAGLCPKCHELVHKDADTAESL TEMKTGLMKKYGGTSVLNQIIPKLVETLADLFPGHFHVTNGWNTKEF REKHHLEKDHDVDAYCIACSHLKPEETLVETEPFEILQFRKHNRAIIHH QTERTYKLDGVTVAKNRKKRMEQKTDSLEDWYVDMAKEHGKTQA DAMRSRLTVIKSTRYYNTPGRMMPGTVFLYEGKRYVMTGQITNGKY YRAYGQEKRNFPAVKVRILTKNTGLVFVA*
SEQ ID 25277 (ωRNA)	GTCAATAACCCATGACTGAAGTCATGGGCTTGCAGATGCAGGTCC TGATGGAAGAAAGGGTTACTGAGCAGAGCAGTGACATGTCATTCG CCGCGGGGTGATTCCAAGCTCCGCGCTCCGGCTAGACATGCCCAT GCTATGGAAACTTTAACGGTATGTGCGGTTTTCCGCTCATACCGGC TTACAACAAATAAGGAGTTATTAG

[0156] Table 1 comprises an example unmodified IscB (SEQ ID 25276) also referred to as RD8_117, and an example ωRNA (SEQ ID 25277). In example embodiments, RD8_117 is utilized as a reference IscB polypeptide to which modifications are made. Modifications of IscB polypeptides can be made by aligning domains and sequences to the reference IscB polypeptide.

[0157] An unmodified IscB polypeptide may be selected from SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145,

147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, 891, 893, 895, 897, 899, 901, 903, 905, 907, 909, 911, 913, 915, 917, 919, 921, 923, 925, 927, 929, 931, 933, 935, 937, 939, 941, 943, 945, 947, 949, 951, 953, 955, 957, 959, 961, 963, 965, 967, 969, 971, 973, 975, 977, 979, 981, 983, 985, 987, 989, 991, 993, 995, 997, 999, 1001, 1003, 1005, 1007, 1009, 1011, 1013, 1015, 1017, 1019, 1021, 1023, 1025, 1027, 1029, 1031, 1033, 1035, 1037, 1039, 1041, 1043, 1045, 1047, 1049, 1051, 1053, 1055, 1057, 1059, 1061, 1063, 1065, 1067, 1069, 1071, 1073, 1075, 1077, 1079, 1081, 1083, 1085, 1087, 1089, 1091, 1093, 1095, 1097, 1099, 1101, 1103, 1105, 1107, 1109, 1111, 1113, 1115, 1117, 1119, 1121, 1123, 1125, 1127, 1129, 1131, 1133, 1135, 1137, 1139, 1141, 1143, 1145, 1147, 1149, 1151, 1153, 1155, 1157, 1159, 1161, 1163, 1165, 1167, 1169, 1171, 1173, 1175, 1177, 1179, 1181, 1183, 1185, 1187, 1189, 1191, 1193, 1195, 1197, 1199, 1201, 1203, 1205, 1207, 1209, 1211, 1213, 1215, 1217, 1219, 1221, 1223, 1225, 1227, 1229, 1231, 1233, 1235, 1237, 1239, 1241, 1243, 1245, 1247, 1249, 1251, 1253, 1255, 1257, 1259, 1261, 1263, 1265, 1267, 1269, 1271, 1273, 1275, 1277, 1279, 1281, 1283, 1285,

1287, 1289, 1291, 1293, 1295, 1297, 1299, 1301, 1303, 1305, 1307, 1309, 1311, 1313, 1315, 1317, 1319, 1321, 1323, 1325, 1327, 1329, 1331, 1333, 1335, 1337, 1339, 1341, 1343, 1345, 1347, 1349, 1351, 1353, 1355, 1357, 1359, 1361, 1363, 1365, 1367, 1369, 1371, 1373, 1375, 1377, 1379, 1381, 1383, 1385, 1387, 1389, 1391, 1393, 1395, 1397, 1399, 1401, 1403, 1405, 1407, 1409, 1411, 1413, 1415, 1417, 1419, 1421, 1423, 1425, 1427, 1429, 1431, 1433, 1435, 1437, 1439, 1441, 1443, 1445, 1447, 1449, 1451, 1453, 1455, 1457, 1459, 1461, 1463, 1465, 1467, 1469, 1471, 1473, 1475, 1477, 1479, 1481, 1483, 1485, 1487, 1489, 1491, 1493, 1495, 1497, 1499, 1501, 1503, 1505, 1507, 1509, 1511, 1513, 1515, 1517, 1519, 1521, 1523, 1525, 1527, 1529, 1531, 1533, 1535, 1537, 1539, 1541, 1543, 1545, 1547, 1549, 1551, 1553, 1555, 1557, 1559, 1561, 1563, 1565, 1567, 1569, 1571, 1573, 1575, 1577, 1579, 1581, 1583, 1585, 1587, 1589, 1591, 1593, 1595, 1597, 1599, 1601, 1603, 1605, 1607, 1609, 1611, 1613, 1615, 1617, 1619, 1621, 1623, 1625, 1627, 1629, 1631, 1633, 1635, 1637, 1639, 1641, 1643, 1645, 1647, 1649, 1651, 1653, 1655, 1657, 1659, 1661, 1663, 1665, 1667, 1669, 1671, 1673, 1675, 1677, 1679, 1681, 1683, 1685, 1687, 1689, 1691, 1693, 1695, 1697, 1699, 1701, 1703, 1705, 1707, 1709, 1711, 1713, 1715, 1717, 1719, 1721, 1723, 1725, 1727, 1729, 1731, 1733, 1735, 1737, 1739, 1741, 1743, 1745, 1747, 1749, 1751, 1753, 1755, 1757, 1759, 1761, 1763, 1765, 1767, 1769, 1771, 1773, 1775, 1777, 1779, 1781, 1783, 1785, 1787, 1789, 1791, 1793, 1794, 1796, 1798, 1800, 1802, 1804, 1806, 1808, 1810, 1812, 1814, 1816, 1818, 1820, 1822, 1824, 1826, 1828, 1830, 1832, 1834, 1836, 1838, 1840, 1842, 1844, 1846, 1848, 1850, 1852, 1854, 1856, 1858, 1860, 1862, 1864, 1866, 1868, 1870, 1872, 1874, 1876, 1878, 1880, 1882, 1884, 1886, 1888, 1890, 1892, 1894, 1896, 1898, 1900, 1902, 1904, 1906, 1908, 1910, 1912, 1914, 1916, 1918, 1920, 1922, 1924, 1926, 1928, 1930, 1932, 1934, 1936, 1938, 1940, 1942, 1944, 1946, 1948, 1950, 1952, 1954, 1956, 1958, 1960, 1962, 1964, 1966, 1968, 1970, 1972, 1974, 1976, 1978, 1980, 1982, 1984, 1986, 1988, 1990, 1992, 1994, 1996, 1998, 2000.

[0158] The IscB polypeptide family of proteins include a larger IscB protein, also referred to as IscB or large IscB, and two smaller IscB proteins, also referred to as IsrB and IshB and detailed further herein. IscB is ~400 amino acids (aa) long and contains a RuvC endonuclease domain split by the insertion of a bridge helix (BH) and an HNH endonuclease domain, an architecture that is shared with Cas9. In one example embodiment, IscB proteins may comprise a N-terminal PLMP domain, a RuvC endonuclease, and a HNH domain. The RuvC domain may be a split RuvC domain comprising RuvC-I, RuvC-II, and RuvC-III subdomains. A bridge helix domain may be inserted between two of the RuvC domains. In one example embodiment,

the bridge helix domain is inserted between the RuvC-I and RuvC-II subdomains. Unlike Cas9, unmodified IscB polypeptides do not contain a Rec domain. IscB proteins may also further comprise a conserved C-terminal domain.

[0159] The IscB polypeptide may comprise an inactive RuvC domain, an inactive HNH domain, or both. In an embodiment, the unmodified IscB polypeptide comprises an inactive RuvC domain. In one embodiment, the IscB polypeptide comprising an inactive RuvC domain is a nickase. In one embodiment, the unmodified IscB polypeptide has a sequence identity of at least 80%, at least 85%, at least 90%, or at least 95% with a wildtype IscB polypeptide, in an embodiment, an IscB sequence selected from SEQ ID NOs: 24525-25062.

[0160] The unmodified IscB polypeptide may comprise an inactive HNH domain; in one embodiment the IscB polypeptide comprising an inactive HNH domain is a nickase. In one embodiment, the IscB nuclease has a sequence identity of at least 80%, at least 85%, at least 90%, or at least 95% with a wildtype IscB polypeptide, in an embodiment an IscB sequence selected from SEQ ID NOS: 2605-23711.

[0161] In an embodiment, the IscB polypeptide comprises an inactive RuvC domain and an inactive HNH domain; in one embodiment the unmodified IscB polypeptide comprises an inactive RuvC domain and an inactive HNH domain and is catalytically inactive. In one embodiment, the IscB polypeptide has a sequence identity of at least 80%, at least 85%, at least 90%, or at least 95% with a wildtype IscB polypeptide, in particular embodiment the IscB sequence is selected from SEQ ID NOs. 25063-25118.

[0162] In an embodiment, the IscB protein is an IsrB (Insertion sequence RuvC-like OrfB), named to emphasize their distinct domain architecture, replacing the previous designation, IscB1 (Kapitonov, V. et al. (2015), *J. Bacteriol.* **198**, 797–807). IsrB polypeptides are shorter, ~350 aa IscB homologs that are also encoded in IS200/605 superfamily transposons. These proteins contain a PLMP domain and split RuvC but lack the HNH domain.

[0163] The unmodified IscB type protein may be an IshB polypeptide (Insertion sequence HNH-like OrfB). IshB are a family of ~180 aa proteins that only contained the PLMP domain and HNH domain but no RuvC domain,

[0164] In one example embodiment, an IscB polypeptide comprises, moving from the N- to C-terminus, a PLMP domain, a RuvC-I subdomain, a bridge helix, a RuvC-II subdomain, a HNH domain, a RuvC-III subdomain, and a C terminal domain. The following sections provide an overview of an example IscB domain architecture.

RuvC domain

[0165] The IscB polypeptides comprise a RuvC domain, which may comprise multiple subdomains, *e.g.*, RuvC-I, RuvC-II and RuvC-III. The subdomains may be separated by interval sequences on the amino acid sequence of the protein.

[0166] Examples of RuvC domains include any polypeptides having a structural similarity and/or sequence similarity to a RuvC domain described in the art. For example, the RuvC domain may share a structural similarity and/or sequence similarity to a RuvC of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 5%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC domains.

[0167] In some examples, the RuvC domain comprise RuvC-I polypeptide, RuvC-II polypeptide, and RuvC-III polypeptide. Examples of the RuvC-I domain also include any polypeptides having a structural similarity and/or sequence similarity to a RuvC-I domain described in the art. For example, the RuvC-I domain may share a structural similarity and/or sequence similarity to a RuvC-I of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 5%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC-I domain. The RuvC-II domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC-II domain described in the art. For example, the RuvC-II domain may share a structural similarity and/or sequence similarity to a RuvC-II of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 5%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC-II domains. The RuvC-III domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC-III domain described in the art. For example, the RuvC-III domains may share a structural similarity and/or sequence similarity to a RuvC-III of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 5%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC-III domains.

[0168] For example, and as described in the art (e.g., Crystal structure of Cas9 in complex with guide RNA and target DNA, Nishimasu et al. Cell, 2014) the RuvC domain of Cas9 consists of a six-stranded mixed β-sheet (β1, β2, β5, β11, β14 and β17) flanked by α-helices

(α33, α34 and α39–α45) and two additional two-stranded antiparallel β-sheets (β3/β4 and β15/β16). It has been described that the RuvC domain of Cas9 shares structural similarity with the retroviral integrase superfamily members characterized by an RNase H fold, such as Escherichia coli RuvC (PDB code 1HJR, 14% identity, root-mean-square deviation (rmsd) of 3.6 Å for 126 equivalent Cα atoms) and Thermus thermophilus RuvC (PDB code 4LD0, 12% identity, rmsd of 3.4 Å for 131 equivalent Cα atoms). E. coli RuvC is a 3-layer alpha-beta sandwich containing a 5-stranded beta-sheet sandwiched between 5 alpha-helices. RuvC nucleases have four catalytic residues (e.g., Asp7, Glu70, His143 and Asp146 in T. thermophilus RuvC), and cleave Holliday junctions (or structurally analogous cruciform junctions) through a two-metal mechanism. Asp10 (Ala), Glu762, His983 and Asp986 of the Cas9 RuvC domain are located at positions similar to those of the catalytic residues of *T. thermophilus* RuvC.

[0169] In an example embodiment, split Ruv-C domain of the IscB proteins may have an HNH domain located between the Ruv-C II and Ruv-C III subdomains as described in more detail below. For example, the IscB protein domain architecture is comprised of the PLMP (P) domain, RuvC-I-II-III domains, a bridge domain (B), an HNH domain and a 3' terminal carboxyl (C) domain spanning 494 amino acids. The bridge domain is located between the RuvC-I and RuvC-II domains and the HNH domain is located between the RuvC-II and RuvC-III domains. A portion thereof may be any smaller (e.g., any 1% of to 95% of the whole) unit (i.e., portion) of a RuvC domain. The portion thereof may be capable of preforming a function of the domain, wherein the function may be the same, increased, or decreased compared to the whole RuvC domain.

HNH domain

[0170] HNH domain comprise two antiparallel β strands connected with a variable length loop, an alpha helix, with a metal binding site between the two. The HNH conserved sites are conserved across the HNH superfamily, with HNH conservation throughout bacteria. In Cas9 proteins, for example, the HNH domain comprises a two-stranded antiparallel β-sheet (β12 and β13) flanked by four α-helices (α35–α38). It shares structural similarity with the HNH endonucleases characterized by a ββα-metal fold, such as phage T4 endonuclease VII (Endo VII) (PDB code 2QNC, 20% identity, rmsd of 2.7 Å for 61 equivalent Cα atoms) and Vibrio vulnificus nuclease (PDB code 1OUP, 8% identity, rmsd of 2.7 Å for 77 equivalent Cα atoms). HNH nucleases have three catalytic residues (e.g., Asp40, His41, and Asn62 in Endo VII), and

cleave nucleic acid substrates through a single-metal mechanism. In the structure of the Endo VII N62D mutant in complex with a Holliday junction, a Mg2+ ion is coordinated by Asp40, Asp62, and the oxygen atoms of the scissile phosphate group of the substrate, while His41 acts as a general base to activate a water molecule for catalysis. Asp839, His840, and Asn863 of the Cas9 HNH domain correspond to Asp40, His41, and Asn62 of Endo VII, respectively, consistent with the observation that His840 is critical for the cleavage of the complementary DNA strand. The N863A mutant functions as a nickase, indicating that Asn863 participates in catalysis. The Cas9 HNH domain may cleave the complementary strand of the target DNA through a single-metal mechanism, as observed for other HNH superfamily nucleases. Although the Cas9 HNH domain shares a ββα-metal fold with other HNH endonucleases, their overall structures are distinct, consistent with the differences in their substrate specificities. Accordingly, IscB polypeptides of the present invention may comprises similar HNH domains in terms of sequence and/or function and may likewise comprise mutations analogous to those described above for Cas9 which convert the IscB polypeptide to a nickase. In an exemplary embodiment, a mutation to catalytic RuvC-II residue corresponding to E157A in corresponding to the sequence numbering of AwaIscB in an IscB polypeptide can be performed to abolish or significantly reduce the nucleolytic activity on the non-target DNA strand.

PLMP Domain

The IscB polypeptides comprise a conserved N-terminal domain, which is referred to herein as a PLMP domain or an X domain. In embodiments, the N-terminal X domain may have one or more conserved residues and/or motifs as identified in FIG. 3 and FIG. 10; see also FIG. 4-3 for PLMP motif alignment. In one embodiment, the PLMP domain comprises a conserved PLMP (SEQ ID NO:2372) amino acid motif. The PLMP motif can be located at or near the N terminus of the IscB polypeptide, including, for example at amino acids 12-15 of AwaIscB, or amino acids corresponding to *A. warmingii* IscB.

[0171] In some examples, the PLMP domain may be no more than 10, no more than 20, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 amino acids in length. For example, the PLMP domain may be no more than 70 amino acids in length, such as comprising 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, or 70 amino acids in length. PLMP domains may be found

upstream of the RuvC-I domain and/or Bridge Helix, where present, of an IscB polypeptide. In one embodiment, the PLMP domain is located within 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 20 or 10 amino acids upstream of the RuvC-1 domain.

[0172] In an aspect, truncation of the N-terminus domain of an IscB polypeptide, including. More than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 amino acids, up to 70 amino acids of the N terminus, i.e. truncation of the PLMP domain, abolishes activity of the IscB polypeptide. In an aspect, more than 4 amino acids PLMP domain may reduce or abolish IscB activity. C-terminal domain

[0173] The C-terminal domain (also referred to herein as a Y domain) may comprise one or more conserved residues or motifs. The C-terminal domain may be no more than 10, no more than 20, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 amino acids in length. For example, the Y domain may be no more than 70 amino acids in length, such as comprising 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, or 70 amino acids in length.

[0174] In an aspect, the IscB polypeptide, comprises a C-terminal domain that is structurally homologous to a tudor domain. See, e.g. Ren et al., Cell Res. (2014) 24:1146-1149. Tudor domains typically comprise a barrel-shaped beta strand fold and range in size around 50 and 60 amino acids. See, e.g. Kawale, A.A. & Burmann, B.M. Inherent backbone dynamics fine-tune the functional plasticity of Tudor domains. Structure (2021), incorporated herein by reference; see, in particular, Figure 1 showing exemplary tudor domain structure.

[0175] In an aspect, the IscB polypeptide, comprises a Tudor Lance (TL) region that are large secondary structure extensions from the tudor domain, see FIGs. 18-20. The TL region may bind to DNA via positive charge contacts. The Lance region of the TL region is oriented in a manner such that extension of the region may result in larger contacts with the DNA major groove near the TAM region of the DNA. The TL region is diverse but typically comprises large hydrophobic amino acids and possesses an overall charge. In general, the TL region may comprise one or more alpha helices and/or one or more beta sheets. The TL region may be ~25 to 50 amino acids long. In example embodiments, the TL regions is 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, 60, 61, 62, 63, 64, 65, 67, 68, 69, 70, 80, or 90 amino acids long.

In an embodiment, the Lance region of the TL region is oriented in a manner such that extension of the region may result in larger contacts with the DNA major groove near the TAM region of the DNA.

Bridge helix

[0176] The nucleic-acid guided nuclease comprises a bridge helix (BH) domain. The bridge helix domain refers to a helix and arginine rich polypeptide. The bridge helix domain may be located next to anyone of the amino acid domains in the nucleic-acid guided nuclease. In one embodiment, the bridge helix domain is next to a RuvC domain, e.g., next to RuvC-I, RuvC-II, or RuvC-III subdomain. In one example, the bridge helix domain is between a RuvC-1 and RuvC2 subdomains.

[0177] The bridge helix domain may be from 10 to 100, from 20 to 60, from 30 to 50, e.g., 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46 or 47, 48, 49, or 50 amino acids in length. Examples of bridge helix includes the polypeptide of amino acids 60-93 of the sequence of S. pyogenes Cas9.

Examples of the BH domain also include any polypeptides a structural similarity and/or sequence similarity to a BH domain described in the art. For example, the BH domain may share a structural similarity and/or sequence similarity to a BH domain of Cas9.

CHIMERIC ISCB POLYPEPTIDES

[0178] Chimeric IscB polypeptides as referenced herein, start with an IscB polypeptide, such as any of those described above, and modify the IscB polypeptide to include one or more heterologous polypeptide sequences to derive a chimeric IscB polypeptide. As used herein "heterologous" refers to polypeptide sequences that are not native to the unmodified IscB and originate from a source other than the IscB polypeptide being modified. The polypeptide sequences may comprise a functional domain, or functional fragment thereof. Heterologous polypeptide sequences may be added directly to a N-terminus or a C-terminus of the IscB polypeptide being modified, inserted within the IscB polypeptide being modified, or a combination thereof. An insertion of a heterologous sequence may occur between existing amino acids anywhere on the IscB polypeptide. An example location may be a junction, for example a loop region or structure, in between domains. A loop region may form between larger domains comprising alpha helices, beta sheets, or both. A loop region may also form between individual alpha helices or beta sheets. A junction may further comprise a conserved

loop region or region with high homology among IscB orthologs. In an example embodiment, one or more insertions of a heterologous polypeptide occurs at a junction.

[0179] In addition to insertions of heterologous polypeptide sequences that do not replace or remove any the existing polypeptide sequence of the IscB polypeptide being modified, substitution where the heterologous polypeptide sequence replaces a portion of the existing polypeptide sequence of the IscB polypeptide being modified are also envisions. In addition to substitutions, some chimeric polypeptides may also require deletion of an existing polypeptide sequences of the IscB being modified. For example, a deletion may be required to remove steric hindrance or other constraints imposed by introduction of the heterologous polypeptide sequence into the IscB polypeptide being modified.

[0180] In one example embodiment, the heterologous polypeptide fragment may be from a homolog or ortholog of IscB. The terms "ortholog" and "homolog" are well known in the art. By means of further guidance, a "homolog" refers to two genes that share a common ancestral gene. Homologous proteins may but need not be structurally related or are only partially structurally related. An "ortholog" are two genes that share common ancestral gene but occur in different species. Orthologous proteins may but need not be structurally related or are only partially structurally related. In one example embodiment, the heterologous polypeptide sequence may comprise a RuvC domain, a HNH domain, a PLMP domain, a TAM interacting (TI) domain, or a RuvC domain from an ortholog or homolog of the IscB polypeptide being modified.

[0181] One or more domains from example IscB polypeptides are utilized herein for insertion including, IscB_Rd4_7, IscB_Rd8_149, ChlorIscB, Iscb_Rd8_127, CRISPR IscB 00644, IscB_large_28, IscB_Rd8_75, IscB_Rd8_151, IscB_Rd8_23, IscB_Rd8_24, and IscB_Rd8_118. Further example IscB polypeptides for use in the chimeric IscB polypeptides are further described herein.

[0182] In one example embodiment, the heterologous polypeptide fragment may be from a CRISPR-Cas system. The heterologous polypeptide sequence or domain may be derived from a a Type 1, a Type II, a Type III, a Type IV, a Type V, or a Type VI CRISPR-Cas system. The heterologous polypeptide sequence may comprise a Rec domain, a RuvC domain, a HNH domain, or a PAM interacting (PI) domain. One or more domains from example CRISPR-Cas polypeptides are utilized herein for insertion including ProCas9-2, Cas9_665, Cas9_1261, and

Cas9_1079. Further example polypeptides for use in the chimeric IscB polypeptides are further described herein.

Modified Rec Domains

[0183] In an example embodiment, a Rec domain, or functional fragment thereof is inserted into an IscB polypeptide. The Rec domain may comprise multiple subdomains, e.g., Rec1, Rec2 and Rec3. The subdomains may be separated by interval sequences on the amino acid sequence of the protein.

[0184] Examples of Rec domains include any polypeptides having a structural similarity and/or sequence similarity to a Rec domain described in the art. For example, the Rec domain may share a structural similarity and/or sequence similarity to a Rec domain of a Type II Cas polypeptide. In another example, the Rec domain may share a structural similarity and/or sequence similarity to a Rec domain of Cas9. In example embodiments, the Rec domain may share a structural similarity and/or sequence similarity to *Francisella novicida* (FnoCas9), *Neisseria meningitidis* (NmeCas9), *Staphylococcus aureus* (SaCas9), *Streptococcus pyogenes* (SpCas9), *Streptococcus thermophilus* (StCas9), *Acidothermus cellulolyticus* (AceCas9), *Campylobacter jejuni* (CjeCas9) or a combination thereof. In some examples, the Rec domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 5%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with Rec domains. Example sequences of insertions of rec domains in Rd8_117 reference IscB polypeptide from Table 1 may comprise a sequence from Figures 27-42, 46-48, 51.

[0185] The crystal structure information (described in International Patent Publication No. Publication of WO 2015089364 A1, 61/980,012 filed April 15, 2014; and Nishimasu *et al*, "Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA," Cell 156(5):935–949, DOI: dx.doi.org/10.1016/j.cell.2014.02.001 (2014), each and all of which are incorporated herein by reference) provides structural information to truncate and create modular or multi-part CRISPR enzymes which may be incorporated into inducible composition. In particular, structural information is provided for *S. pyogenes* Cas9 (SpCas9), and this may be extrapolated to other Cas9 orthologs or IscB proteins (as well as homologs and orthologs thereof) or other nucleic acid-guided nucleases. In one embodiment, the conformational variations in the crystal structures of the CRISPR-Cas9 system or of components of the CRISPR-Cas9 provide important and critical information about the flexibility or movement of protein structure

regions relative to nucleotide (RNA or DNA) structure regions that may be important for the function of other IscB polypeptide and related systems. The structural information provided for Cas9 (e.g., *S. pyogenes* Cas9) may be used to further engineer and optimize use in IscB polypeptides as well as interrogate structure-function relationships of related Cas systems.

[0186] In some examples, the Rec domain comprise Rec1 polypeptide, Rec2 polypeptide, and/or Rec3 polypeptide. Examples of the Rec1 domain also include any polypeptides having a structural similarity and/or sequence similarity to a Rec1 domain described in the art. For example, the Rec1 domain may share a structural similarity and/or sequence similarity to a Rec1 of Cas9. In some examples, the Rec domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with Rec1 domain. The Rec2 domain also include any polypeptides of structural similarity and/or sequence similarity to a Rec2 domain described in the art. For example, the Rec2 domain may share a structural similarity and/or sequence similarity to a Rec2 of Cas9. In some examples, the Rec domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with Rec2 domains. The Rec3 domain also include any polypeptides of structural similarity and/or sequence similarity to a Rec3 domain described in the art. For example, the Rec3 domains may share a structural similarity and/or sequence similarity to a Rec3 of Cas9. In some examples, the Rec domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with Rec3 domains.

[0187] For example, and as described in the art (e.g., Crystal structure of Cas9 in complex with guide RNA and target DNA, Nishimasu et al. Cell, 2014) the Rec domain of Cas9 consists of three regions, a long α helix referred to as the bridge helix, the REC1 domain, and the REC2 domain. REC1 adopts an elongated, α-helical structure comprising 25 α helices (α2–α5 and α12–α32) and two β sheets (β6 and β10 and β7–β9), whereas REC2 adopts a six-helix bundle structure (α6–α11).

[0188] In an example embodiment, the Rec domain is inserted at a junction in the IscB polypeptide or analogous position of another IscB polypeptide. In an example embodiment, the

Rec domain is inserted between amino acids 153-160 of RD8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide.

Nucleotide Binding Proteins

[0189] The IscB polypeptide may include a nucleotide binding domain, which may comprise a protein or fragment thereof, to enhance or extend the functionality of the IscB polypeptide. A nucleotide binding domain may enhance functionality by stabilizing one or more domains of the IscB polypeptide, for example the polynucleotide components of the IscB. The functionality of the IscB polypeptide may be extended by the nucleotide binding domain by adding additional nuclease activity or additional substrate binding.

[0190] In an example embodiment, a nucleotide binding domain is inserted into the IscB polypeptide. The nucleotide binding domain may be capable of binding to RNA, DNA, or both. In general, such a system may comprise a nuclease (e.g., an endonuclease and/or exonuclease), a hybrid binding domain, or a RuvC domain associated (e.g., fused) with a IscB polypeptide, e.g., IscB protein. In an example embodiment, the nucleotide binding domain is inserted at a junction in the IscB polypeptide or analogous position of another IscB polypeptide. In an example embodiment, the nucleotide binding domain is inserted between amino acids 153-160 of amino acids 153-160 of RD8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide.

Endo/Exonucleases

[0191] In an example embodiment, an endonuclease, or fragment thereof, is inserted into the IscB polypeptide. Endonucleases, or restriction endonucleases, bind and cleave internal strands of nucleic acids. Endonucleases may bind RNA and DNA. Endonucleases are classified into different types according to their structure, recognition site, cleavage site, cofactor(s), and activator(s). These Endonuclease types are I, II, III, and IV which further include subclasses. Depending on the endonuclease, one or more subunits and one or more holoenzymes may be required to form the restriction, methylase, and specificity domain. An endonuclease may be selected for stabilizing the IscB polypeptide, targeting a substrate, or conferring an additional function. These domains may be associated (e.g. fused) on one or more IscB polypeptides.

[0192] Endonucleases requires the presence of Mg^{2+} for nuclease activity and S-adenosyl methionine for methylase stimulation or activity. In general, the recognition sites are 4, 5, 6, 7, or 8 bases long as well as palindromic. In an example embodiment, the endonuclease is a Type II endonuclease. Type II endonucleases are, in general, homodimers around 25 and 35 kDa per

monomer. The homodimers typically form a 3D "U" shaped dimeric holoenzyme wherein the recognition domains form the sides and bridging domains at the bottom. Type II share a common core comprising five β-sheets flanked on each side by an α-helix. Some endonucleases are capable of binding to and hydrolyzing ssDNA or ssRNA.

[0193] Fragments of the endonuclease may be fused to IscB polypeptides. In example embodiments, the recognition domain is fused to an IscB polypeptides. This may comprise various spacers and constructs. Various endonucleases and fusion sites are known in the art and may be utilized herein. See Williams, Raymond J. "Restriction endonuclease." Molecular biotechnology 23.3 (2003): 225-243. In an example embodiment, an endonuclease or fragment thereof is associated (e.g., fused) with a junction in an IscB polypeptide or analogous position of another IscB polypeptide. In an example embodiment, the endonuclease, or fragment thereof, is inserted between around amino acids 153-160 amino acids 153-160 of RD8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide.

[0194] In an examine embodiment, an exonuclease, or fragment thereof, is inserted into the IscB polypeptide. Exonucleases bind and cleave the 3' or 5' ends of nucleic acids. Exonucleases may bind to RNA and DNA. Exonucleases are classified into different types according to their function. These types are I, II, III, IV, V, VII, and Lambda.

RNase/DNase

[0195] In an example embodiment, an RNase, or fragment thereof, is inserted into the IscB polypeptide. A ribonuclease (RNase) is a polypeptide that binds to and hydrolyzes RNA substrates. Natural RNases can be found in both bacterial and eukaryotic enzymes and are well known in the art. RNases are typically characterized by the substrates they bind and ribonucleolytic activity. The RNase, or fragment thereof, may comprise a RNA binding domain.

[0196] The IscB polypeptide may include am RNase, or fragment thereof, to enhance or extend the functionality of the IscB polypeptide. An RNase may enhance functionality by stabilizing one or more domains of the IscB polypeptide, for example the polynucleotide components of the IscB. The functionality of the IscB polypeptide may be extended by an RNase by adding additional nuclease activity or additional substrate binding. An RNase may be associated (e.g., fused) with an IscB polypeptide. This association may be at a junction (further described herein). An RNase may be associated at a junction at or near the surface or

the IscB polypeptide or a junction wherein the association does not result in steric clashes between the IscB polypeptide and the RNase.

[0197] In an example embodiment, the RNase is an endoribonuclease or exoribonuclease. In an example embodiment, the endoribonuclease comprises RNase A, RNase H, RNase H, RNase III, RNase L, RNase P, RNase PhyM, RNase T1, RNase T2, RNase US, RNase V, RNase E, or RNase G. In an example embodiment, the exoribonuclease comprises exoribonuclease I, exoribonuclease II, RNase PH, RNase R, RNase D, RNase T.

[0198] In an example embodiment, an RNase or fragment thereof is associated (e.g., fused) with a junction in an IscB polypeptide or analogous position of another IscB polypeptide. In an example embodiment, an RNase, or fragment thereof, is inserted between around amino acids 153-160 of amino acids 153-160 of RD8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide. Example RNase fusions that may be applied to IscB polypeptides by one skilled in the art can be found in Table 1 of Gotte, G.; Menegazzi, M. Biological Activities of Secretory RNases: Focus on Their Oligomerization to Design Antitumor Drugs. Frontiers in Immunology, 2019, 10 incorporated herein by reference.

[0199] In an example embodiment, a DNase, or fragment thereof, is inserted into the IscB polypeptide. A deoxyribonuclease (DNase) is a polypeptide that binds to and hydrolyzes DNA substrates and are well known in the art. The DNase, or fragment thereof, may comprise a DNA binding domain. In general, DNases are categorized into two families by their biochemical and biological properties: DNase I, which comprises of DNase 1L1, DNase 1L2, and DNase 1L3, and DNase II, which comprises DNase II α, DNase II β and L-DNase II. DNase I require Mg^{2+} and Ca^{2+} for catalytic activity while DNase II does not. In an example embodiment, the DNase is an endodeoxyribonuclease or an exodeoxyribonuclease.

[0200] In an example embodiment, the DNase is inserted at a junction in the IscB polypeptide or analogous position of another IscB polypeptide. In an example embodiment, the DNase, or fragment thereof, is inserted between amino acids around 153-160 of amino acids 153-160 of RD8_117 polypeptide of Table 1 or an analogous position of another IscB polypeptide. See Lauková, L.; Konečná, B.; Janovičová, Ľ.; Vlková, B.; Celec, P. Deoxyribonucleases and Their Applications in Biomedicine. Biomolecules, 2020, 10, 1036.

Hybrid Binding Domain

[0201] In an example embodiment, a hybrid binding domain (HBD), or functional fragment thereof, is inserted into the IscB polypeptide. The HBD confers binding to RNA/DNA

hybrids as well as binding to dsRNA and dsDNA. The HBD may comprise a domain at an amino terminal (N-terminal) of a RNase I. The HBD domain of RNase I is composed of a three-stranded antiparallel β sheet and two short helices. In general, HBD has two separate regions that independently bind to RNA and DNA. A protein loop in the HBD recognizes the RNA strand and forms hydrogen bonds with the 2'-OH groups while polar residues in the HBD interact with the phosphate groups and aromatic residues are selective for deoxyriboses. See Nowotny, Marcin et al. "Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD." The EMBO journal vol. 27,7 (2008): 1172-81. doi:10.1038/emboj.2008.44 and Wang, K.; Wang, H.; Li, C.; Yin, Z.; Xiao, R.; Li, Q.; Xiang, Y.; Wang, W.; Huang, J.; Chen, L.; Fang, P.; Liang, K. Genomic Profiling of Native R Loops with a DNA-RNA Hybrid Recognition Sensor. Science Advances, 2021, 7.

[0202] The IscB polypeptide may include an HBD, or fragment thereof, to enhance or extend the functionality of the IscB polypeptide. An HBD may enhance functionality by stabilizing one or more domains of the IscB polypeptide, for example the polynucleotide components of the IscB. The functionality of the IscB polypeptide may be extended by an HBD by adding additional nuclease activity or additional substrate binding. An HBD may be associated (e.g., fused) with an IscB polypeptide. This association may be at a junction (further described herein). An HBD may be associated at a junction at or near the surface or the IscB polypeptide or a junction wherein the association does not result in steric clashes between the IscB polypeptide and the HBD. In an example embodiment, the HDB is inserted at a junction in the IscB polypeptide or analogous position of another IscB polypeptide. In an example embodiment, the HBD is inserted between amino acids around 153-160 of amino acids 153-160 of RD8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide.

RuvC domains

[0203] In an example embodiment, a RuvC domain, or fragment thereof, is inserted into the IscB polypeptide. The RuvC domain may comprise multiple subdomains, e.g., RuvC-I, RuvC-II and RuvC-III. The subdomains may be separated by interval sequences on the amino acid sequence of the protein.

[0204] The IscB polypeptide may include a RuvC domain, or fragment thereof, to enhance or extend the functionality of the IscB polypeptide. A RuvC domain may enhance functionality by stabilizing one or more domains of the IscB polypeptide, for example the polynucleotide components of the IscB. The functionality of the IscB polypeptide may be extended by an

RuvC domain by adding additional nuclease activity or additional substrate binding. An RuvC domain may be associated (e.g. fused) with an IscB polypeptide. This association may be at a junction (further described herein). A RuvC domain may be associated at a junction at or near the surface or the IscB polypeptide or a junction wherein the association does not result in steric clashes between the IscB polypeptide and the RuvC domain.

[0205] Examples of the RuvC domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC domain described in the art. For example, the RuvC domain may share a structural similarity and/or sequence similarity to a RuvC of Cas9.

[0206] In some examples, the RuvC domain comprise RuvC-I polypeptide, RuvC-II polypeptide, and RuvC-III polypeptide. Examples of the RuvC-I domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC-I domain described in the art. For example, the RuvC-I domain may share a structural similarity and/or sequence similarity to a RuvC-I of Cas9. The RuvC-II domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC-II domain described in the art. For example, the RuvC-II domain may share a structural similarity and/or sequence similarity to a RuvC-II of Cas9. The RuvC-III domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC-III domain described in the art. For example, the RuvC-III domains may share a structural similarity and/or sequence similarity to a RuvC-III of Cas9.

[0207] For example, and as described in the art (e.g. Crystal structure of Cas9 in complex with guide RNA and target DNA, Nishimasu et al. Cell, 2014) the RuvC domain of Cas9 consists of a six-stranded mixed β-sheet (β1, β2, β5, β11, β14 and β17) flanked by α-helices (α33, α34 and α39–α45) and two additional two-stranded antiparallel β-sheets (β3/β4 and β15/β16). It has been described that the RuvC domain of Cas9 shares structural similarity with the retroviral integrase superfamily members characterized by an RNase H fold, such as Escherichia coli RuvC (PDB code 1HJR, 14% identity, root-mean-square deviation (rmsd) of 3.6 Å for 126 equivalent Cα atoms) and Thermus thermophilus RuvC (PDB code 4LD0, 12% identity, rmsd of 3.4 Å for 131 equivalent Cα atoms). RuvC nucleases have four catalytic residues (e.g., Asp7, Glu70, His143 and Asp146 in T. thermophilus RuvC), and cleave Holliday junctions through a two-metal mechanism. Asp10 (Ala), Glu762, His983 and Asp986 of the Cas9 RuvC domain are located at positions similar to those of the catalytic residues of T. thermophilus RuvC. There are key structural discrepancies between the Cas9 RuvC domain and the RuvC nucleases, which explain their functional differences. Unlike the Cas9 RuvC

domain, the RuvC nucleases form dimers and recognize Holliday junctions. In addition to the conserved RNase H fold, the Cas9 RuvC domain has other structural elements involved in interactions with the guide:target heteroduplex (an end-capping loop between α42 and α43) and the PI domain/stem loop 3 (β-hairpin formed by β3 and β4).

[0208] In one embodiment, the nucleic acid-guided nuclease comprises at least one nuclease domain. In an embodiment, the nucleic acid-guided nuclease protein comprises at least two nuclease domains. In an embodiment, the one or more nuclease domains are only active upon presence of a cofactor. In an embodiment, the cofactor is Magnesium (Mg). In embodiments where more than one nuclease domain is present and the substrate is a double-strand polynucleotide, the nuclease domains each cleave a different strand of the double-strand polynucleotide. In an embodiment, the nuclease domain is a RuvC domain. In an example embodiment the RuvC domain lacks nuclease activity.

[0209] In an example embodiment, the RuvC is inserted at a junction in the IscB polypeptide or analogous position of another IscB polypeptide. In an example embodiment, the RuvC domain, or fragment thereof, is inserted between amino acids around 153-160 of amino acids 153-160 of RD8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide.

TAM/PAM Interaction Domain Modifications

[0210] The IscB polypeptides may comprise modifications to, or insertions of, a TAM interacting (TI) domain from another IscB or Omega system, a WED/adaptor stabilizer domain from another IscB or Omega system, or a PAM Interacting (PI) domain from a CRISPR-Cas system. In one embodiment, insertion preserves RNA interaction with the TAM determining region, or improves RNA interaction with the TAM determining region. The TI domain insertion or WED/adaptor stabilizer insertion can be from another IscB polypeptide and allows for the TAM to be modified relative to the wild-type IscB polypeptide.

[0211] The IscB systems disclosed may recognize a target adjacent motif (TAM) in order to recognize and bind a target sequence on a target polynucleotide. In one embodiment, the IscB polypeptide and related compositions do not contain a TAM requirement or may be engineered to comprise a TI domain or functional fragment thereof. The precise sequence and length requirements for the TAM will differ depending on the IscB polypeptide used, and/or the TI domain insertion. In an example embodiment, TAM determining region comprises one or more amino acid substitutions.

[0212] In some examples, TAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). In one example embodiment, the TAM is 3' adjacent to the target polynucleotide. In another example embodiment, the TAM is 5' adjacent to the target sequence of the target polynucleotide. In one embodiment, the cleavage site is distant from the Target Adjacent Motif (TAM), *e.g.*, the cleavage occurs after the nth nucleotide on the non-target strand and after the nucleotide on the targeted strand. In one embodiment, the cleavage site occurs after an identified nucleotide (counted from the TAM) on the non-target strand and after the further identified nucleotide (counted from the TAM) on the targeted strand.

[0213] The WED/adaptor stabilizer region, is also referred to the wedge (WED) domain, see FIG. 9. WED domains are typically identified as oligonucleotide binding domains and may aid in recognition and/or stability of RNA scaffolds, and have intermolecular interactions with the RNA scaffold structure. The WED domain may be adjacent to or in proximity to a TAM interacting domain in an unmodified IscB polypeptide. In an embodiment, the WED domain may have structural similarity to the WED domain of Cas9, which comprises a fold with a twisted five-stranded beta sheet flanked by four alpha helices and is responsible for the recognition of the distorted repeat: anti-repeat duplex. See, Morlot et al., *J. Biol. Chem.* 280, 15984-91, (2005). The WED domain of an unmodified IscB polypeptide may comprise one or more antiparallel β sheet or β strands flanked by one or more alpha helices.

[0214] In an example embodiment, the TI domain, WED domain, PI domain, or functional fragment thereof is inserted between amino acids 365-499 of IscB polypeptide Rd8_117 from Table 1, or an analogous position of another IscB polypeptide. In an example embodiment, the TI domain, WED domain, PI domain, or functional fragment thereof is from a Type II Cas polypeptide. In an example embodiment, the TI domain or functional fragment thereof is from an IscB polypeptide. In an example embodiment, the TAM of the wild-type IscB polypeptide is retained or wherein the TAM is modified. In an example embodiment, the insertion replaces amino acids 365-499, 369-499, 462-486 (Lance), 450-486 (Tudor), or 376-383, 436-448, and 462-486 (Lance_TIL) of the TI domain of Rd8_117 of Table 1, or an analogous position of another IscB polypeptide.

[0215] In an example embodiment, the insertion comprises 380-735 from SEQ ID NO: 2365, one or more amino acid positions 556-609 from cA2 ProCas9-2, one or more amino acid positions 356-420 from IscB_Rd8_149, one or more amino acid positions 386-464 from

IscB_Rd4_7, one or more amino acid positions 856-924 from Cas9_971, one or more amino acid positions 569-751 from Cas9_1079_3, one or more amino acid positions from ChlorIscB, one or more amino acid positions 488-512 from SEQ ID NO: 2367, one or more amino acid positions 739-765 from SEQ ID NO 2365, one or more amino acid positions 407-431 from IscB_Rd8_127, one or more amino acid positions 404-430 from CRISPR IscB 00644, one or more amino acid positions 376-482 from IscB_large_28, one or more amino acid positions 356-488 from IscB_Rd8_149, one or more amino acid positions 376_482 from IscB_Rd8_75, one or more amino acid positions 374-495 from IscB_Rd8_151, one or more amino acid positions 376-477 from IscB_Rd8_23, one or more amino acid positions 376-477 from IscB_Rd8_24, one or more amino acid positions 376-486 from IscB_Rd8_118, one or more amino acid positions 374-495 from IscB_Rd8_151, and/or one or more amino acid positions 376-486 amino acids from IscB_Rd8_118,In an example embodiment, any one or more of the TI, Tudor, or Tudor Lance domain insertions are performed on an analogous position in the WED domain with an analogues domain. Example sequences of insertions or substitutions in Rd8_117 reference IscB polypeptide from Table 1 may comprise a sequence from any one of Figures 52 to 95.

[0216] In one example embodiment, the IscB polypeptides lack or substantially lack a PAM interacting (PI) domain. In an embodiment, the IscB polypeptides may be engineered to comprise a PI domain or a functional fragment of a PI domain. In an embodiment, the IscB polypeptides may achieve a target specificity by a non-protein domain. In an embodiment, targeting specificity is obtained by a central hairpin structure in a guide molecule.

[0217] Examples of PAM sequences for the IscB polypeptides herein include NGG and NAC. For example, the IscB polypeptide may recognize PAM sequence NAC. In an example embodiment, wherein the domain is an NGG PI or TI domain, or functional fragment thereof.

[0218] The PAM interaction domain or PI domain as referred to herein is reported to be responsible for determining PAM specificity of IscB polypeptides. By means of example, the PI domain is contained in the NUC lobe and forms an elongated structure comprising seven α-helices, a three-stranded antiparallel β-sheet, a five-stranded antiparallel β-sheet, and a two-stranded antiparallel β-sheet.

[0219] In some cases, where the IscB polypeptide do have a PAM requirement, the precise sequence and length requirements for the PAM will differ depending on the IscB polypeptide used. In some examples, PAMs are typically 2-5 base pair sequences adjacent the protospacer

(that is, the target sequence). Examples of the natural PAM sequences for different nucleic acid-guided nucleases orthologs have been identified and the skilled person will be able to identify further PAM sequences for use in the engineered IscB polypeptides.

[0220] Further, associating a PAM Interacting (PI) domain (e.g., attaching or fusing) to a nucleic acid-guided nuclease may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the IscB, genome engineering platform. IscB polypeptide may be engineered to alter their PAM specificity, for example as described in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592. The skilled person will understand that other IscB proteins may be modified analogously.

C-Terminus Modifications

[0221] The C-Terminal domain of the IscB polypeptide may be modified to enhance or extend functionality of the IscB polypeptide. The C-Terminal domain may be enhanced or extended by removing, substituting, or adding one or more amino acids or functional domains to the C-Terminal domain. These modifications (e.g., removing, substituting, or adding) may enhance the IscB polypeptide by, for example, stabilizing the IscB polypeptide or increasing the specificity of target binding. These modifications may also extend the functionality of the IscB polypeptide by, for example, the addition of non-native enzymatic function or promoting on-off switching as further described herein.

[0222] The C-terminal domain may comprise one or more conserved residues or motifs. The C-terminal domain may be no more than 10, no more than 20, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 amino acids in length. For example, the C-terminal domain may be no more than 70 amino acids in length, such as comprising 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, or 70 amino acids in length.

[0223] In an aspect, the IscB polypeptide, comprises a C-terminal domain that is structurally homologous to a tudor domain. See, e.g. Ren et al., Cell Res. (2014) 24:1146-1149. Tudor domains typically comprise a barrel-shaped beta strand fold and range in size around 50 and 60 amino acids. See, e.g. Kawale, A.A. & Burmann, B.M. Inherent backbone

dynamics fine-tune the functional plasticity of Tudor domains. Structure (2021), incorporated herein by reference; see, in particular, Figure 1 showing exemplary tudor domain structure.

[0224] The C-terminus of the present invention may be engineered to comprise additional functionality. Any one or more amino acids of the C-terminus may be substituted to fuse a functional domain to the C-terminus. The one or more amino acids may comprise amino acids that interact with non-target nucleic acid, amino acids that do not interact with non-target nucleic acid, or both.

[0225] In an example embodiment, a C-terminus is engineered for nucleotide editing. In an example a nucleotide deaminase is fused at the C-terminus of the IscB polypeptide. In an example embodiment, the C-terminus is engineered to comprise a transposase. In an example embodiment, the C-terminus is engineered to comprise a base editing system. In an example embodiment, the C-terminus is engineered for a prime editing system. Further examples and descriptions are further described elsewhere herein.

Additional Modifications

[0226] In an example embodiment, one or more amino acids on the surface of the engineered IscB are substituted or removed. Additional modifications may comprise substituting or removing one or more amino acids (i.e., modified amino acids) on the surface of the IscB polypeptide. In an example embodiment, the one or more amino acids on the surface of the engineered IscB are substituted to have a different charge. The one or more charged amino acids on the surface of the IscB polypeptide may be substituted to increase or decrease the surface charge of the IscB polypeptide. The one or more amino acids substituted or removed may carry a charge (e.g., Arginine, Histidine, Lysine, Aspartic Acid, Glutamic Acid). In an example embodiment, the one or more amino acids on the surface of the engineered IscB are removed or substituted to increase or decrease the overall charge of the protein surface. For example, one or more positively charged amino acids (e.g., Arg, His, Lys) are substituted with a negatively charged amino acid (e.g., Asp, Glu). In an example embodiment, wherein the surface charge of the engineered IscB is more neutral relative to the wildtype. For example, one or more positively charged amino acids and one or more negatively charged amino acids are substituted with a neutral charge amino acid.

[0227] The IscB polypeptides may comprise one or more amino acid mutations. The present disclosure provides a mutated IscB polypeptide, having one or more mutations resulting in improved activity, e.g., improved IscB enzymes for use in effecting modifications to target

loci when complexed to ωRNAs, relative to an unmodified or wild-type IscB. The present disclosure provides a mutated chimeric IscB polypeptide, having one or more mutations resulting in improved activity, e.g., improved chimeric IscB enzymes for use in effecting modifications to target loci when complexed to ωRNAs, relative to chimeric IscB without the one or more amino acid modifications. It is to be understood that mutated polypeptides as described herein below may be used in any of the methods according to the present disclosure as described herein elsewhere and with chimeric IscB polypeptides described herein. Any of the methods, products, compositions and uses as described herein elsewhere are equally applicable with the mutated IscB polypeptides as further detailed below. Mutations may be varied in the precise amino acid substitution utilized as described elsewhere herein and can be varied at similar positions to achieve similar effects.

[0228] In one example embodiment, the one or more amino acid mutations increase activity by 1-fold, 1.5-fold, 2.0-fold or more relative to an unmodified (e.g. unmutated or wild-type) IscB polypeptide. In an example embodiment, the IscB polypeptide is mutated at amino acid position H271, Q60, W425, A288, E308, P405, E409, D413, G416, L291, N66, I85, L470, T561, E576, E137, M500, E501, T358, S352, V490, G489, T484, I533, N566, T481, T536, and/or Y538 corresponding to the amino acid positions in of IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide is mutated at an amino acid position identified in Table 3. In an example embodiment, the IscB polypeptide is mutated by extending the DNA/RNA duplex channel and may comprise one or more mutations correspond to E308R, P405R, E409R, D413K, G416K and/or L291R of scB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide is mutated by hydrophobic repacking of RuvC and may comprise one or more mutations corresponding to H271I, Q60I, W425R and A288V of IscB Rd8_117 Cas9 1070. In an example embodiment, the IscB polypeptide is mutated by creating a new or modified non-target strand channel and may comprise one or more mutations correspond to I85R, L470R and/or N66H of IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide is mutated by altering unwinding activity and may comprise one or more mutations correspond to T561R and/or E576R of IscB Rd8_117 Cas9 1079_1 protein. In an example embodiment, the IscB polypeptide is mutated by enhancing the DNA/RNA channel and may comprise one or more mutations correspond to E137K, T358R, S352H, M500K and/or E501K of IscB Rd8_117 Cas9 1079_1 protein of IscB Rd8_117 Cas9

1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide is mutated to alter non-specific DNA binding, and may comprise one or more mutations that correspond to V490K, G489K, T484K, I533K, N566R, T481R, T536R, and/or Y538R of IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). Further example mutations are detailed at Table 3 of the working examples.

[0229] In an example embodiment, the IscB polypeptide mutations comprise Q60, W425, and A288 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise Q60I, W425R and A288V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise W425, and A288 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise Q60I and A288V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise Q60 and W425 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the mutations comprise Q60I and W425R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278).

[0230] In an example embodiment, the IscB polypeptide mutations comprise P405, E409. E308, and L291 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise P405R, E409R, E308R, and L291R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise E409, E308, and L291 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the mutations comprise E409R, E308R, and L291R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the mutations comprise P405, E308, and L291 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise P405R, E308R, and L291R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise P405, E409. and L291 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise P405R, E409R, and L291R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise P405, E409. E308, and L291

corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise P405, E409, and E308 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the mutations comprise P405R, E409R, and E308R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278).

[0231] In an example embodiment, the IscB polypeptide mutations comprise I85, E576, E137, I533, E409, and A88 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise I85R, E576R, E137K, I533K, E409R, and A88V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the mutations comprise E576, E137, I533, E409, and A88 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise E576R, E137K, I533K, E409R, and A88V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise I85, E137, I533, E409, and A88 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise I85R, E137K, I533K, E409R, and A88V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise I85, E576, I533, E409, and A88 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise I85R, E576R, I533K, E409R, and A88V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise I85, E576, E137, E409, and A88 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise I85R, E576R, E137K, E409R, and A88V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise I85, E576, E137, I533, and A88 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise I85R, E576R, E137K, I533K, and A88V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise I85, E576, E137, I533, and E409 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one

embodiment, the IscB polypeptide mutations comprise I85R, E576R, E137K, I533K, and E409R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278).

[0232] In an example embodiment, the IscB polypeptide mutations comprise E409 and E576 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the mutations comprise E409R and E576R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise E409 and A288 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise E409R and A288V corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise E409 and I533 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise E409R and I533K corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise E409 and I85 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise E409R and I85R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the mutations comprise E409 and E137 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise E409R and E137K corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise P405 and E576 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise P405R and E576R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise T561 and E576 corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In one embodiment, the IscB polypeptide mutations comprise T561R and E576R corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). In an example embodiment, the IscB polypeptide mutations comprise E137 and S352. In one embodiment, the IscB polypeptide mutations comprise E137K and S352H corresponding to IscB Rd8_117 Cas9 1079_1 protein (SEQ ID NO: 25278). Further description of the combination mutations is provided in the working examples at Table 4.

[0233] The term “corresponding amino acid” or “residue which corresponds to” refers to a particular amino acid or analogue thereof in an IscB homolog or ortholog that is identical,

functionally or otherwise equivalent to an amino acid in a reference IscB protein. Such equivalency may be based on structural domain or subdomains of the IscB protein and may be according to alignments as described elsewhere herein. Accordingly, as used herein, referral to an "amino acid position corresponding to amino acid position [X]" of a specified IscB protein represents referral to a collection of equivalent positions in other recognized IscB and structural homologues and families. In one example embodiment, mutations are referenced to residues which correspond to the unmodified IscB polypeptide of Table 1. In one example embodiment, mutations are referenced to residues which correspond to the Rd8_117 1079_1 polypeptide of SEQ ID NO: 25278.

[0234] Further modifications may comprise inserting a functional domain, further described herein, in the WED/adaptor stabilizer region, the Tudor or Tudor Lance domain, or combination thereof.

Example Chimeric IscB Systems

[0235] Chimeric IscB systems may be designed using the guidance provided herein using an unmodified IscB system, such as provided in Table 1 to identify suitable positions (e.g., junctions, loops) for substitutions or insertions described herein. Positions for insertions are detailed herein including for the Rd8_117 system of Table 1 and may also be used to align to other IscB polypeptides and ωRNAs to identify analogous positions, junctions, and/or loops for example. Example insertions or substitutions of domains or portions thereof are provided in sequence maps of Figures 25-95, detailing insertion locations and size of insertions and domain replacements.

[0236] An example chimeric IscB polypeptide comprises the Rd8_117 IscB polypeptide with a REC domain insertion from Cas9 1079_1, which may also be referred to as Rd8_117_1079_1. See, FIG. 32. In an embodiment, the Rd8_117 with REC insertion from Cas9 1079_1 comprises the sequence:

MRYVYVLDVDGKPLMPTCRFGKVRRMLKSGQAKAVDTLPFTIQLTYKPRTRILQP
VTLGQDPGRTNIGMAAVRFDGKELGRFHCITRNKEIPKLMADRMAARKASRRGER
LARKRLARKLHTTAKHLNGRILPGCSEPIAVKDIINTESRFNNRILTKCKVCGKN
TPLRRNVRELLLENIVRFLPLESELKETLKRTILEGQQGNINKLFRKLKFNQKDW
PGKNLTDIAKNKLPGRLPFCKEHFAENEKFTTIEKSTFRLTPTATQLLRTHINLF
RKLSGILPVTDVAVELNKFAFMQLDNPEMKKREIDFCHGPLCGTGGLEAAVKEQQ

DGKCLLCGKESIGHYHHIVPRSRRGSNTIANIAGLCPKCHELVHKDADTAESLTE MKTGLMKKYGGTSVLNQIIPKLVETLADLFPGHFHVTNGWNTKEFREKHHLEKDH DVDAYCIACSHLKPEETLVETEPFEILQFRKHNRAIIHHQTERTYKLDGVTVAKN RKKRMEQKTDSLEDWYVDMAKEHGKTQADAMRSRLTVIKSTRYYNTPGRMMPGTV FLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKVRILTKNTGLVFVA (SEQ ID NO: 25278).

ωRNA Molecules and Modifications

[0237] The systems herein may further comprise one or more ωRNA molecules, which are referred to herein interchangeably as ωRNA. The ωRNA complex can comprise a guide sequence and a scaffold that interacts with the IscB polypeptide. An ωRNA molecule may form a complex with IscB polypeptide nuclease or IscB polypeptide and direct the complex to bind with a target sequence. In certain example embodiments, the ωRNA molecule is a single molecule comprising a scaffold sequence and a spacer sequence. In certain example embodiments, the spacer is 5' of the scaffold sequence. In certain example embodiments, the ωRNA molecule may further comprise a conserved nucleic acid sequence between the scaffold and spacer portions.

[0238] In certain example embodiments, the ωRNA scaffold comprises a spacer sequence and a conserved nucleotide sequence. The ωRNA scaffold typically comprises conserved regions, with the scaffold comprising 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 105, 115, 125, 135, 145, 155, 165, 175, 185, 195, 205, 215, 225, 235, 245, 255, 265, 275, 285, 295, 305, 315, 325, 335, 345, or 355 or more nt. In an aspect, the ωRNA scaffold comprises one conserved nucleotide sequence. In embodiments, the conserved nucleotide sequence is on or near a 5' end of the scaffold. In embodiments, the scaffold may comprise a short 3-4 base pair nexus, a conserved nexus hairpin and a large multi-stem loop region that may consist of two interconnected multi-stem loops. In an aspect, an IscrB associated scaffold may comprise a spacer, which can be re-programmed to direct site-specific binding to a target sequence of a target polynucleotide. The spacer may also be referred to herein as part of the ωRNA scaffold or as gRNA and may comprise an engineered heterologous sequence. In an embodiment the scaffold may comprise a sequence from Table 1.

[0239] In an embodiment, the spacer length of the ωRNA is from 10 to 150 nt. In an embodiment, the spacer length of the ωRNA is at least 15 nucleotides. In an embodiment, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27 to 30 nt, e.g., 27, 28, 29, or 30 nt, from 30 to 35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer. In certain example embodiment, the spacer sequence is 15, 16, 17,18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 40, 41, 42, 43, 44, 45, 46, 47 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 17, 138, 19, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149 or 150 nt.

[0240] In an embodiment, the ωRNA spacer length is from 15 to 50 nt. In an embodiment, the spacer length of the ωRNA is at least 15 nucleotides. In an embodiment, the spacer length is from 15 to 50 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27 to 30 nt, e.g., 27, 28, 29, or 30 nt, from 30 to 35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt, from 34 to 40 nt, e.g., 34, 35, 36, 37, 38, 39, 40, from 35 to 39, from 36 to 38 nt long, about 37 nt, or longer.

[0241] In one embodiment, the sequence of the ωRNA molecule is selected to reduce the degree of secondary structure within the ωRNA molecule or may be further engineered to reduce degree of secondary structure within one or more regions of the ωRNA molecule. In one embodiment, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting ωRNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example of a folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A.R. Gruber

et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0242] In an example embodiment, the ωRNA molecule is from the same as the IscB polypeptide. In an example embodiment, the IscB polypeptide is Rd8_117 from Table 1 and the ωRNA molecule is from Rd8_117 of Table 1. In an example embodiment, the ωRNA molecule is optimized for Rd8_117 and comprises the sequence: GTCAATAACCCATGACTGAAGTCATGGGCTTGCAGATGCAGGTCCTGATGGAAG AAAGGGTTACTGAGCAGAGCAGTGACATGTCATTCGCCGCGGGGTGATTCCAAG CTCCGCGCTCCGGCTAGACATGCCCATGCTATGGAAACTTTAACGGTATGTGCGG TTTTCCGCTCATACCGGCTT (SEQ ID NO: 25279). In an embodiment, mutations to the ωRNA molecule for the IscB Rd8_117 polypeptide are made relative to SEQ ID NO: 25279, or to ωRNA from Table 1 (SEQ ID NO: 25277).

[0243] In certain example embodiments, the ωRNA molecule may have a sequence identity of at least 80%, at least 85%, at least 90%, 91%, 92%, 93%, 94%, or at least 95%, 96%, 97%, 98% 99% with a sequence selected from SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688,

690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, 892, 894, 896, 898, 900, 902, 904, 906, 908, 910, 912, 914, 916, 918, 920, 922, 924, 926, 928, 930, 932, 934, 936, 938, 940, 942, 944, 946, 948, 950, 952, 954, 956, 958, 960, 962, 964, 966, 968, 970, 972, 974, 976, 978, 980, 982, 984, 986, 988, 990, 992, 994, 996, 998, 1000, 1002, 1004, 1006, 1008, 1010, 1012, 1014, 1016, 1018, 1020, 1022, 1024, 1026, 1028, 1030, 1032, 1034, 1036, 1038, 1040, 1042, 1044, 1046, 1048, 1050, 1052, 1054, 1056, 1058, 1060, 1062, 1064, 1066, 1068, 1070, 1072, 1074, 1076, 1078, 1080, 1082, 1084, 1086, 1088, 1090, 1092, 1094, 1096, 1098, 1100, 1102, 1104, 1106, 1108, 1110, 1112, 1114, 1116, 1118, 1120, 1122, 1124, 1126, 1128, 1130, 1132, 1134, 1136, 1138, 1140, 1142, 1144, 1146, 1148, 1150, 1152, 1154, 1156, 1158, 1160, 1162, 1164, 1166, 1168, 1170, 1172, 1174, 1176, 1178, 1180, 1182, 1184, 1186, 1188, 1190, 1192, 1194, 1196, 1198, 1200, 1202, 1204, 1206, 1208, 1210, 1212, 1214, 1216, 1218, 1220, 1222, 1224, 1226, 1228, 1230, 1232, 1234, 1236, 1238, 1240, 1242, 1244, 1246, 1248, 1250, 1252, 1254, 1256, 1258, 1260, 1262, 1264, 1266, 1268, 1270, 1272, 1274, 1276, 1278, 1280, 1282, 1284, 1286, 1288, 1290, 1292, 1294, 1296, 1298, 1300, 1302, 1304, 1306, 1308, 1310, 1312, 1314, 1316, 1318, 1320, 1322, 1324, 1326, 1328, 1330, 1332, 1334, 1336, 1338, 1340, 1342, 1344, 1346, 1348, 1350, 1352, 1354, 1356, 1358, 1360, 1362, 1364, 1366, 1368, 1370, 1372, 1374, 1376, 1378, 1380, 1382, 1384, 1386, 1388, 1390, 1392, 1394, 1396, 1398, 1400, 1402, 1404, 1406, 1408, 1410, 1412, 1414, 1416, 1418, 1420, 1422, 1424, 1426, 1428, 1430, 1432, 1434, 1436, 1438, 1440, 1442, 1444, 1446, 1448, 1450, 1452, 1454, 1456, 1458, 1460, 1462, 1464, 1466, 1468, 1470, 1472, 1474, 1476, 1478, 1480, 1482, 1484, 1486, 1488, 1490, 1492, 1494, 1496, 1498, 1500, 1502, 1504, 1506, 1508, 1510, 1512, 1514, 1516, 1518, 1520, 1522, 1524, 1526, 1528, 1530, 1532, 1534, 1536, 1538, 1540, 1542, 1544, 1546, 1548, 1550, 1552, 1554, 1556, 1558, 1560, 1562, 1564, 1566, 1568, 1570, 1572, 1574, 1576, 1578, 1580, 1582, 1584, 1586, 1588, 1590, 1592, 1594, 1596, 1598, 1600, 1602, 1604, 1606, 1608, 1610, 1612, 1614, 1616, 1618, 1620, 1622, 1624, 1626, 1628, 1630, 1632, 1634, 1636, 1638, 1640, 1642, 1644, 1646, 1648, 1650, 1652, 1654, 1656, 1658, 1660, 1662, 1664, 1666, 1668, 1670, 1672, 1674, 1676, 1678, 1680, 1682, 1684, 1686, 1688, 1690, 1692, 1694, 1696, 1698, 1700, 1702, 1704, 1706, 1708, 1710, 1712, 1714,

1716, 1718, 1720, 1722, 1724, 1726, 1728, 1730, 1732, 1734, 1736, 1738, 1740, 1742, 1744, 1746, 1748, 1750, 1752, 1754, 1756, 1758, 1760, 1762, 1764, 1766, 1768, 1770, 1772, 1774, 1776, 1778, 1780, 1782, 1784, 1786, 1788, 1790, 1792, 1795, 1797, 1799, 1801, 1803, 1805, 1807, 1809, 1811, 1813, 1815, 1817, 1819, 1821, 1823, 1825, 1827, 1829, 1831, 1833, 1835, 1837, 1839, 1841, 1843, 1845, 1847, 1849, 1851, 1853, 1855, 1857, 1859, 1861, 1863, 1865, 1867, 1869, 1871, 1873, 1875, 1877, 1879, 1881, 1883, 1885, 1887, 1889, 1891, 1893, 1895, 1897, 1899, 1901, 1903, 1905, 1907, 1909, 1911, 1913, 1915, 1917, 1919, 1921, 1923, 1925, 1927, 1929, 1931, 1933, 1935, 1937, 1939, 1941, 1943, 1945, 1947, 1949, 1951, 1953, 1955, 1957, 1959, 1961, 1963, 1965, 1967, 1969, 1971, 1973, 1975, 1977, 1979, 1981, 1983, 1985, 1987, 1989, 1991, 1993, 1995, 1997, 1999, 2001.

[0244] In an example embodiment, the ωRNA is engineered to reduce steric interactions between the ωRNA and inserted domain, for example, the PK-loop of ωRNA and an inserted REC domain. In an example embodiment, one or more amino acids are modified or inserted thereby reducing the interaction between an inserted domain and the ωRNA. In an example embodiment, one or more nucleotides in a loop, described herein, of the ωRNA is modified and/or inserted. In an example embodiment, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the ωRNA is modified. In an example embodiment, a pair of nucleotides are inserted into the ωRNA. In an example embodiment, at least 20, at least 15, at least 10, at least 5, at least 1 amino acid is inserted into the ωRNA. In an example embodiment, no more than 20, no more than 15, no more than 10, no more 5, no more than 1 amino acid is inserted into the ωRNA.

[0245] Modifications to the ωRNA can occur at various locations along the molecule. In an example embodiment, truncations from the 3' end of the ωRNA molecule of about 5, 10, 15, 20, 25 or 30 nucleotides can be truncated from the 3' end relative to a WT ωRNA. One or more insertions in the nexus pseudoknot can comprise one or more insertions in the nexus stem, which may be upstream or downstream of the insertion in the nexus pseudoknot. Further modifications, including to the pseduoknot and engineering the terminal region as a transRNA on-switch and PK region as a transRNA off-switch are detailed further herein.

[0246] As used herein, a heterologous ωRNA molecule is an ωRNA molecule that is not derived from the same species as the IscB polypeptide nuclease, or comprises a portion of the molecule, e.g., spacer, that is not derived from the same species as the IscB polypeptide, e.g. IscB protein. For example, a heterologous ωRNA molecule of a IscB polypeptide nuclease

derived from species A comprises a polynucleotide derived from a species different from species A, or an artificial polynucleotide.

[0247] In an example embodiment, the ωRNA comprises a peptide nucleic acid (PNA). A PNA may be a synthetic mimic of DNA or RNA. A PNA may comprise a a pseudo-peptide polymer backbone in place of the deoxyribose or ribose phosphate backbone. PNAs are well known in the art. See Pellestor, F., Paulasova, P. The peptide nucleic acids (PNAs), powerful tools for molecular genetics and cytogenetics. Eur J Hum Genet 12, 694–700 (2004) and Abhishek Singhal, Valentina Bagnacani, Roberto Corradini, and Peter E. Nielsen ACS Chemical Biology 2014 9 (11), 2612-2620. In an example embodiment, a single strand or both strands of the ωRNA comprises a PNA. In an example embodiment, a portion of one or both strands comprise a PNA. In an example embodiment, no more than 75%, no more than 50%, no more than 40%, no more than 30%, no more than 25%, no more than 20%, no more than 15%, no more than 10%, or no more than 5%, of the one or both strands of the ωRNA comprise a PNA. In an example embodiment, no less than 75% of one or both strands of the ωRNA comprise a PNA.

[0248] In a particular embodiment, the ωRNA comprises a guide sequence linked to a conserved nucleotide sequence, wherein the conserved nucleotide sequence may comprise one or more stem loops or optimized secondary structures. In an embodiment, the conserved nucleotide sequence has a minimum length of 16 nts and a single stem loop. In further embodiments the conserved nucleotide sequence has a length longer than 16 nts, preferably more than 17 nts, and has more than one stem loops or optimized secondary structures. In one embodiment, the guide sequence may be linked to all or part of the natural conserved nucleotide sequence. In one embodiment, certain aspects of the guide architecture can be modified, for example by addition, subtraction, or substitution of features, whereas certain other aspects of guide architecture are maintained. Preferred locations for engineered guide modifications, including but not limited to insertions, deletions, and substitutions include guide termini and regions of the guide that are exposed when complexed with IscB polypeptide nuclease and/or target, for example the tetraloop and/or loop2.

[0249] In one embodiment, a loop in the guide RNA is provided. This may be a stem loop or a tetra loop. The loop is preferably GAAA, but it is not limited to this sequence or indeed to being only 4bp in length. Indeed, preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer

or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG.

[0250] In one embodiment, the ωRNA forms a stem-loop with a separate non-covalently linked sequence, which can be DNA or RNA. In an embodiment, the sequences forming the guide are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). In one embodiment, these sequences can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sufonyl, ally, propargyl, diene, alkyne, and azide. Once this sequence is functionalized, a covalent chemical bond or linkage can be formed between this sequence and the conserved nucleotide sequence. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, sulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0251] In one embodiment, these stem-loop forming sequences can be chemically synthesized. In one embodiment, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0252] The repeat:anti repeat duplex will be apparent from the secondary structure of the ωRNA. It may be typically a first complimentary stretch after (in 5' to 3' direction) the poly U tract and before the tetraloop; and a second complimentary stretch after (in 5' to 3' direction) the tetraloop and before the poly A tract. The first complimentary stretch (the "repeat") is complimentary to the second complimentary stretch (the "anti-repeat"). As such, they Watson-

Crick base pair to form a duplex of dsRNA when folded back on one another. As such, the anti-repeat sequence is the complimentary sequence of the repeat and in terms to A-U or C-G base pairing, but also in terms of the fact that the anti-repeat is in the reverse orientation due to the tetraloop.

[0253] In an embodiment of the invention, modification of guide architecture comprises replacing bases in stem-loop 2. For example, in one embodiment, "actt" ("acuu" in RNA) and "aagt" ("aagu" in RNA) bases in stem-loop 2 are replaced with "cgcc" and "gcgg". In one embodiment, "actt" and "aagt" bases in stem-loop 2 are replaced with complimentary GC-rich regions of 4 nucleotides. In one embodiment, the complimentary GC-rich regions of 4 nucleotides are "cgcc" and "gcgg" (both in 5' to 3' direction). In one embodiment, the complimentary GC-rich regions of 4 nucleotides are "gcgg" and "cgcc" (both in 5' to 3' direction). Other combination of C and G in the complimentary GC-rich regions of 4 nucleotides will be apparent including CCCC and GGGG.

[0254] In one aspect, the stem-loop 2, e.g., "ACTTgtttAAGT" (SEQ ID NO: 1) can be replaced by any "XXXXgtttYYYY" (SEQ ID NO: 2), e.g., where XXXX and YYYY represent any complementary sets of nucleotides that together will base pair to each other to create a stem.

[0255] As used herein, the term "spacer" may also be referred to as a "guide sequence." In one embodiment, the degree of complementarity of the guide sequence to a given target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. In certain example embodiments, the ωRNA molecule comprises a guide sequence that may be designed to have at least one mismatch with the target sequence, such that an RNA duplex formed between the sequence and the target sequence. Accordingly, the degree of complementarity is less than 99%. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less. In one embodiment, the guide sequence is designed to have a stretch of two or more adjacent mismatching nucleotides, such that the degree of complementarity over the entire sequence is further reduced. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less, more particularly, about 92% or less, more particularly about 88% or less, more particularly about 84% or less, more particularly about 80% or less, more particularly about 76% or less, more particularly about 72% or less, depending on whether the

stretch of two or more mismatching nucleotides encompasses 2, 3, 4, 5, 6 or 7 nucleotides, etc. In one embodiment, aside from the stretch of one or more mismatching nucleotides, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a sequence (within a nucleic acid-targeting guide sequence) to direct sequence-specific binding of a nucleic acid -targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a ωRNA system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence (or a sequence in the vicinity thereof) may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the sequence to be tested and a control sequence different from the test guide sequence, and comparing binding or rate of cleavage at or in the vicinity of the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting ωRNA may be selected to target any target nucleic acid sequence.

[0256] A ωRNA sequence, and hence a nucleic acid-targeting guide, may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In one embodiment, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small

cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0257] In one embodiment, the ωRNA molecule forms a stem-loop with a separate non-covalently linked sequence, which can be DNA or RNA. In one embodiment, the sequences forming the ωRNA are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). In one embodiment, these sequences can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sufonyl, ally, propargyl, diene, alkyne, and azide. Once this sequence is functionalized, a covalent chemical bond or linkage can be formed between this sequence and the conserved nucleotide sequence. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, sulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0258] In one embodiment, these stem-loop forming sequences can be chemically synthesized. In one embodiment, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

Pseudoknot Modifications

[0259] The ωRNA may comprise one or more pseudoknot structures. A pseudoknot may be located between the loop of one hairpin in the multi-stem loop region and the region directly downstream of the nexus. This pseudoknot is termed the nexus pseudoknot hairpin (Figs. 4, 15). It is has been demonstrated mutating the pseudoknot sequence while maintaining base pairing had no effect on cleavage activity. However, scrambling the sequence in the pseudoknot and disrupting the pseudoknot structure disrupted the cleavage of a chimeric IscB system. These results suggested that the pseudoknot structure, but not necessarily sequence, was important for ncRNA function. The nexus pseudoknot was present in all major iscB- associated ωRNAs. Another conserved pseudoknot is located between the guide adapter and a hairpin loop in the multi-stem loop region. In some instances a third pseudoknot is formed by a third hairpin in the multi-stem loop region. See Altae-Tran, H.; Kannan, S.; Demircioglu, F. E.; Oshiro, R.; Nety, S. P.; McKay, L. J.; Dlakić, M.; Inskeep, W. P.; Makarova, K. S.; Macrae, R. K.; Koonin, E. V.; Zhang, F. The Widespread IS200/IS605 Transposon Family Encodes Diverse Programmable RNA-Guided Endonucleases. Science, 2021, 374, 57–65.

[0260] The ωRNA may additionally comprise a hairpin that encompasses a Shine-Dalgarno (SD) sequence located approximately 10 bp upstream of the CDS start codon, implying that the ωRNA might be involved in the regulation of the translation of IscB.

[0261] In an example embodiment, the pseudoknot comprises a peptide nucleic acid (PNA). In an example embodiment, the 3'-end of the ωRNA is truncated. In an example embodiment, the 3'-end of the ωRNA is truncated by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides.

Engineered On-Off Switching

[0262] In an example embodiment, the ωRNA is engineered to comprise an on-off switch. An on-off switch is a mechanism wherein, for example, a molecule is introduced to the IscB such that the IscB gains or loses activity, the activity can comprise any function described herein. For example, an engineered IscB may be engineered to be modify a substrate however, the IscB cannot modify the substrate until an "On" molecule is introduced. Thus, the IscB is off in absence of the "On" molecule. Therefore, an ωRNA comprises an on-off switch if, for example, another molecule must bind to the ωRNA to confer or prohibit activity. For example, ωRNA on-off engineering can be conferred by modifying (e.g. substituting) mismatched nucleotides in a ωRNA hairpin thus controlling nuclease activity.

[0263] In an example embodiment, one or more nucleotides in a pseudoknot nexus of the ωRNA is modified and/or inserted. The one or more nucleotides modified or inserted may confer the ωRNA with an on-off switch. The one or more nucleotides modified or inserted may or may not base pair with a corresponding nucleotide but may maintain the structure of the ωRNA. In an example embodiment, one or more nucleotides in a nexus stem of the ωRNA that base pair to the one or more nucleotides in the pseudoknot nexus is modified and/or inserted. In an example embodiment, a base pair comprising a nucleotide in the pseudoknot nexus and a nucleotide in the nexus stem is substituted with a complementary base pair. In an example embodiment, one or more nucleotides in a pseudoknot region of the ωRNA are modified and/or inserted and wherein the nexus stem retains base pairing. In an example embodiment, one or more nucleotides in a pseudoknot region of the ωRNA are modified and/or inserted and wherein the pseudoknot retains its structure relative to a wild-type IscB. In an example embodiment, the pseudoknot region of the ωRNA retains its structure by no less than 50%, no less than 55%, no less than 60%, no less than 65%, no less than 70%, no less than 75%, no less than 80%, no less than 85%, no less than 90%, no less than 95% relative to a wild-type IscB

[0264] An on-off switch may also be conferred to the IscB polypeptide by truncating the 3'-end of the ωRNA and then re-introducing the truncated 3'-end in trans, e.g., transRNA. The trans 3'-end may supplied independently of the ωRNA. In an example embodiment, the 3'-end of the ωRNA is truncated. In an example embodiment, the 3'-end of the ωRNA is truncated by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides. In an example embodiment, the ωRNA is truncated by no more 5, 10, 15, 20, or 25 nucleotides. In example embodiments, any one or more positions from around 151-167 of the ωRNA of Rd8_117 from Table 1, or an analogous position thereof, is truncated and supplied in trans as the transRNA.

[0265] In an example, if positions 151-167 is truncated then a similar sequence as 168-184 is supplied in trans thereby activating nuclease activity. In an example embodiment, an RNA supplied in trans is bound to a 3'-end of the ωRNA. In an example embodiment, the trans RNA is native to the target location, or molecule targeted by the engineered IscB location. In an example embodiment, the transRNA is reprogrammed to target other sequences. In an example embodiment, the transRNA can be increased or decreased. In an example embodiment, the transRNA is increased or decreased by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides. In an example embodiment, the transRNA is increased or

decreased by no more than 5, 10, 15, 20, or 25 nucleotides. Such a transRNA on-switch can enable cell-type specific or temporally controlled editing for added layers of specificity.

[0266] In an example embodiment, a complimentary nucleotide sequence is introduced to the IscB polypeptide. For example, a sequence complimentary to an accessible region of the ωRNA is introduced to the IscB polypeptide. The complimentary sequence may bind to the accessible region of the ωRNA resulting in disrupted nuclease activity. In an example embodiment, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides are targeted by the complimentary sequence. In an example embodiment, no more than 15, 14, 13, 12, 11, 10, or 5 nucleotides are targeted by the complementary sequence. In an example embodiment, an accessible region of the ωRNA is a hairpin external to the polypeptide. In an example embodiment, an accessible region of the ωRNA is the nexus pseudoknot. In an example embodiment, the accessible region is positions around and between 100-107 or 139-150 of Rd8_117 ωRNA from Table 1 or an analogous position thereof.

Other ωRNA Modifications

Chemical Modifications

[0267] In an embodiment, the ωRNA molecule comprises non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Preferably, these non-naturally occurring nucleic acids and non-naturally occurring nucleotides are located outside the ωRNA sequence. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a ωRNA nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a ωRNA comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the ωRNA comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, a locked nucleic acid (LNA) nucleotide comprising a methylene bridge between the 2′ and 4′ carbons of the ribose ring, or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, or 2'-fluoro analogs. Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine, inosine, 7-methylguanosine. Examples of ωRNA chemical modifications include, without limitation, incorporation of 2′-O-methyl (M), 2′-O-methyl 3′phosphorothioate (MS), S-

constrained ethyl(cEt), or 2′-O-methyl 3′thioPACE (MSP) at one or more terminal nucleotides. Such chemically modified ωRNA can comprise increased stability and increased activity as compared to unmodified ωRNA, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015 Ragdarm et al., 0215, PNAS, E7110-E7111; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Li et al., Nature Biomedical Engineering, 2017, 1, 0066 DOI:10.1038/s41551-017-0066). In one embodiment, the 5' and/or 3' end of a ωRNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al., 2016, J. Biotech. 233:74-83). In an embodiment, a ωRNA comprises ribonucleotides in a region that binds to a target sequence and one or more deoxyribonucleotides and/or nucleotide analogs in a region that binds to the IscB polypeptide nuclease. In an embodiment, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered ωRNA structures. In one embodiment, 3-5 nucleotides at either the 3' or the 5' end of a ωRNA is chemically modified. In one embodiment, only minor modifications are introduced in the seed region, such as 2'-F modifications. In one embodiment, 2'-F modification is introduced at the 3' end of a ωRNA. In an embodiment, three to five nucleotides at the 5' and/or the 3' end of the ωRNA are chemically modified with 2'-O-methyl (M), 2'-O-methyl 3' phosphorothioate (MS), S-constrained ethyl(cEt), or 2'-O-methyl 3' thioPACE (MSP). Such modification can enhance genome editing efficiency (see Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989). In an embodiment, all of the phosphodiester bonds of a ωRNA are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In an embodiment, more than five nucleotides at the 5' and/or the 3' end of the ωRNA are chemically modified with 2'-O-Me, 2'-F or S-constrained ethyl(cEt). Such chemically modified ωRNA can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, PNAS, E7110-E7111). In an embodiment of the invention, a ωRNA is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), or Rhodamine. In certain embodiment, the chemical moiety is conjugated to the ωRNA by a linker, such as an alkyl chain. In an embodiment, the chemical moiety of the modified ωRNA can be used to

attach the ωRNA to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified ωRNA can be used to identify or enrich cells generically edited by a IscB polypeptide nuclease and related systems (see Lee et al., eLife, 2017, 6:e25312, DOI:10.7554).

[0268] In a particular embodiment, the conserved nucleotide sequence may be modified to comprise one or more protein-binding RNA aptamers. In a particular embodiment, one or more aptamers may be included such as part of optimized secondary structure. Such aptamers may be capable of binding a bacteriophage coat protein as detailed further herein.

[0269] In embodiments, the IscB polypeptide utilizes the ωRNA scaffold comprising a polynucleotide sequence that facilitates the interaction with the IscB protein, allowing for sequence specific binding and/or targeting of the guide sequence with the target polynucleotide. Chemical synthesis of the ωRNA scaffold is contemplated, using covalent linkage using various bioconjugation reactions, loops, bridges, and non-nucleotide links via modifications of sugar, internucleotide phosphodiester bonds, purine and pyrimidine residues. Sletten et al., Angew. Chem. Int. Ed. (2009) 48:6974-6998; Manoharan, M. Curr. Opin. Chem. Biol. (2004) 8: 570-9; Behlke et al., Oligonucleotides (2008) 18: 305-19; Watts, et al., Drug. Discov. Today (2008) 13: 842-55; Shukla, et al., ChemMedChem (2010) 5: 328-49; chemical synthesis using automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0270] In certain example embodiments, the scaffold and spacer may be designed as two separate molecules that can hybridize or covalently joined into a single molecule. Covalent linkage can be via a linker (e.g., a non-nucleotide loop) that comprises a moiety such as spacers, attachments, bioconjugates, chromophores, reporter groups, dye labeled RNAs, and non-naturally occurring nucleotide analogues. More specifically, suitable spacers for purposes of this invention include, but are not limited to, polyethers (e.g., polyethylene glycols, polyalcohols, polypropylene glycol or mixtures of ethylene and propylene glycols), polyamines group (e.g., spennine, spermidine and polymeric derivatives thereof), polyesters (e.g., poly(ethyl acrylate)), polyphosphodiesters, alkylenes, and combinations thereof. Suitable attachments include any moiety that can be added to the linker to add additional properties to the linker, such as but not limited to, fluorescent labels. Suitable bioconjugates include, but are not limited to, peptides, glycosides, lipids, cholesterol, phospholipids, diacyl glycerols and

dialkyl glycerols, fatty acids, hydrocarbons, enzyme substrates, steroids, biotin, digoxigenin, carbohydrates, polysaccharides. Suitable chromophores, reporter groups, and dye-labeled RNAs include, but are not limited to, fluorescent dyes such as fluorescein and rhodamine, chemiluminescent, electrochemiluminescent, and bioluminescent marker compounds. The design of example linkers conjugating two RNA components are also described in WO 2004/015075.

[0271] The linker (e.g., a non-nucleotide loop) can be of any length. In one embodiment, the linker has a length equivalent to about 0-16 nucleotides. In one embodiment, the linker has a length equivalent to about 0-8 nucleotides. In one embodiment, the linker has a length equivalent to about 0-4 nucleotides. In one embodiment, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in International Patent Publication No. WO 2011/008730.

Escorted ωRNA molecules

[0272] In one embodiment, the compositions or complexes have a ωRNA molecule with a functional structure designed to improve ωRNA molecule structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[0273] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." Nature Reviews Drug Discovery 9.7 (2010): 537-550). These characteristics also suggest a wide range of uses for aptamers as drug delivery vehicles (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." Trends in biotechnology 26.8 (2008): 442-449; and Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." J Clin Invest 2000, 106:923-928.). Aptamers may also be constructed that function as molecular switches, responding to a que by changing properties, such as RNA aptamers that bind fluorophores to mimic the activity of green fluorescent protein (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." Science 333.6042

(2011): 642-646). It has also been suggested that aptamers may be used as components of targeted siRNA therapeutic delivery systems, for example targeting cell surface proteins (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." Silence 1.1 (2010): 4).

[0274] Accordingly, in one embodiment, the ωRNA molecule is modified, e.g., by one or more aptamer(s) designed to improve ωRNA molecule delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the ωRNA molecule deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends a ωRNA molecule that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O2 concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g., ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[0275] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents. Crytochrome-2 activation is also highly sensitive, allowing for the use of low light intensity stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[0276] Energy sources such as electromagnetic radiation, sound energy or thermal energy may induce the guide. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm2. In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[0277] The chemical or energy sensitive ωRNA may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a ωRNA and have the IscB polypeptide nuclease system or complex function. The invention can involve applying the chemical source or energy so as to have the ωRNA function and the IscB polypeptide nuclease system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[0278] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisic Acid (ABA) (see, e.g., stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. FKBP-FRB based system inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[0279] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., www.pnas.org/content/104/3/1027.abstract). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[0280] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel-based system inducible by energy, heat or radio-wave (see, e.g., www.sciencemag.org/content/336/6081/604). These TRP family proteins respond to different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the ωRNA and the other components of the IscB polypeptide nuclease/ωRNA molecule complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the

nucleus, the ωRNA protein and the other components of the IscB polypeptide nuclease/ωRNA molecule complex will be active and modulating target gene expression in cells.

[0281] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for *in vivo* applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[0282] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under *in vivo* conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 μs and 500 milliseconds, preferably between 1 μs and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[0283] As used herein, 'electric field energy' is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under *in vivo* conditions (see WO97/49450).

[0284] As used herein, the term "electric field" includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell. Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc., as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[0285] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[0286] Electroporation has been used in both *in vitro* and *in vivo* procedures to introduce foreign material into living cells. With *in vitro* applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform *in vitro* electroporation include the Electro Cell Manipulator ECM600 product, and

the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[0287] The known electroporation techniques (both *in vitro* and *in vivo*) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 .mu.s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[0288] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vitro* conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under *in vitro* conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vivo* conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[0289] Preferably, the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term “pulse” includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[0290] Preferably, the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[0291] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue, or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[0292] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm2 to about 100 W/cm2. Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[0293] As used herein, the term “ultrasound” refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[0294] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm2 (FDA recommendation), although energy densities of up to 750 mW/cm2 have been used. In physiotherapy, ultrasound is typically used as an energy source in a range up to about 3 to 4 W/cm2 (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm2 (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic, and focused ultrasound.

[0295] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142. Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106.

[0296] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will appreciate that any variety of combinations of ultrasound may be used. Additionally, the energy density, frequency of ultrasound, and period of exposure may be varied.

[0297] Preferably, the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm-2. Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm-2.

[0298] Preferably, the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source

is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[0299] Preferably the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[0300] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm-2 to about 10 Wcm-2 with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm-2, but for reduced periods of time, for example, 1000 Wcm-2 for periods in the millisecond range or less.

[0301] Preferably, the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[0302] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm-2 or 1.25 Wcm-2 as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[0303] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues unlike light. It is therefore better suited to whole-tissue penetration (such as but not limited to a lobe of the liver) or whole organ (such as but not limited to the entire liver or an entire muscle, such as the heart) therapy. Another important advantage is that ultrasound is a non-invasive stimulus which is used in a wide variety of diagnostic and therapeutic applications. By way of example, ultrasound is well known in medical imaging techniques and, additionally, in orthopedic therapy. Furthermore, instruments suitable for the application of ultrasound to a subject vertebrate are widely available and their use is well known in the art.

[0304] In one embodiment, the ωRNA molecule is modified by a secondary structure to increase the specificity of the IscB polypeptide nuclease and related system and the secondary structure can protect against exonuclease activity and allow for 5' additions to the ωRNA sequence also referred to herein as a protected ωRNA molecule.

[0305] In one aspect, the invention provides for hybridizing a "protector RNA" to a sequence of the ωRNA molecule, wherein the "protector RNA" is an RNA strand complementary to the 3' end of the ωRNA molecule to thereby generate a partially double-stranded ωRNA. In an embodiment of the invention, protecting mismatched bases (i.e., the bases of the ωRNA molecule which do not form part of the ωRNA sequence) with a perfectly complementary protector sequence decreases the likelihood of target DNA binding to the mismatched base pairs at the 3' end. In one embodiment of the invention, additional sequences comprising an extended length may also be present within the ωRNA molecule such that the ωRNA comprises a protector sequence within the ωRNA molecule. This "protector sequence" ensures that the ωRNA molecule comprises a "protected sequence" in addition to an "exposed sequence" (comprising the part of the ωRNA sequence hybridizing to the target sequence). In one embodiment, the ωRNA molecule is modified by the presence of the protector ωRNA to comprise a secondary structure such as a hairpin. Advantageously there are three or four to thirty or more, e.g., about 10 or more, contiguous base pairs having complementarity to the protected sequence, the ωRNA sequence or both. It is advantageous that the protected portion does not impede thermodynamics of the IscB polypeptide nuclease and related system interacting with its target. By providing such an extension including a partially double stranded ωRNA molecule, the ωRNA molecule is considered protected and results in improved specific binding of the IscB polypeptide nuclease/ωRNA molecule complex, while maintaining specific activity.

[0306] In one embodiment, use is made of a truncated ωRNA (tru-ωRNA), i.e., a ωRNA molecule which comprises a ωRNA sequence which is truncated in length with respect to the canonical ωRNA sequence length. As described by Nowak et al. (Nucleic Acids Res (2016) 44 (20): 9555-9564), such guides may allow catalytically active IscB polypeptide nuclease to bind its target without cleaving the target DNA. In one embodiment, a truncated ωRNA is used which allows the binding of the target but retains only nickase activity of the IscB polypeptide nuclease.

[0307] In one embodiment, conjugation of triantennary N-acetyl galactosamine (GalNAc) to oligonucleotide components may be used to improve delivery, for example delivery to select cell types, for example hepatocytes (see International Patent Publication No. WO 2014/118272 incorporated herein by reference; Nair, JK et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961). This is considered to be a sugar-based particle and further details on other particle delivery systems and/or formulations are provided herein. GalNAc can therefore be considered to be a particle in the sense of the other particles described herein, such that general uses and other considerations, for instance delivery of said particles, apply to GalNAc particles as well. A solution-phase conjugation strategy may for example be used to attach triantennary GalNAc clusters (mol. wt. ~2000) activated as PFP (pentafluorophenyl) esters onto 5′-hexylamino modified oligonucleotides (5′-HA ASOs, mol. wt. ~8000 Da; Østergaard et al., Bioconjugate Chem., 2015, 26 (8), pp 1451–1455). Similarly, poly(acrylate) polymers have been described for *in vivo* nucleic acid delivery (see WO2013158141 incorporated herein by reference). In further alternative embodiments, pre-mixing IscB polypeptide nuclease nanoparticles (or protein complexes) with naturally occurring serum proteins may be used in order to improve delivery (Akinc A et al, 2010, Molecular Therapy vol. 18 no. 7, 1357–1364).

[0308] Screening techniques are available to identify delivery enhancers, for example by screening chemical libraries (Gilleron J. et al., 2015, Nucl. Acids Res. 43 (16): 7984-8001). Approaches have also been described for assessing the efficiency of delivery vehicles, such as lipid nanoparticles, which may be employed to identify effective delivery vehicles for components (see Sahay G. et al., 2013, Nature Biotechnology 31, 653–658).

HDR Donor Templates

[0001] In one embodiment, the compositions and systems herein may further comprise one or more nucleic acid templates. In some cases, the nucleic acid template may comprise one or more polynucleotides. In certain cases, the nucleic acid template may comprise coding sequences for one or more polynucleotides. The nucleic acid template may be a DNA template.

[0002] The donor polynucleotide may be used for editing the target polynucleotide. In some cases, the donor polynucleotide comprises one or more mutations to be introduced into the target polynucleotide. Examples of such mutations include substitutions, deletions, insertions, or a combination thereof. The mutations may cause a shift in an open reading frame on the target polynucleotide. In some cases, the donor polynucleotide alters a stop codon in the

target polynucleotide. For example, the donor polynucleotide may correct a premature stop codon. The correction may be achieved by deleting the stop codon or introduces one or more mutations to the stop codon. In other example embodiments, the donor polynucleotide addresses loss of function mutations, deletions, or translocations that may occur, for example, in certain disease contexts by inserting or restoring a functional copy of a gene, or functional fragment thereof, or a functional regulatory sequence or functional fragment of a regulatory sequence. A functional fragment refers to less than the entire copy of a gene by providing sufficient nucleotide sequence to restore the functionality of a wild type gene or non-coding regulatory sequence (e.g., sequences encoding long non-coding RNA). In certain example embodiments, the systems disclosed herein may be used to replace a single allele of a defective gene or defective fragment thereof. In another example embodiment, the systems disclosed herein may be used to replace both alleles of a defective gene or defective gene fragment. A "defective gene" or "defective gene fragment" is a gene or portion of a gene that when expressed fails to generate a functioning protein or non-coding RNA with functionality of the corresponding wild-type gene. In certain example embodiments, these defective genes may be associated with one or more disease phenotypes. In certain example embodiments, the defective gene or gene fragment is not replaced but the systems described herein are used to insert donor polynucleotides that encode gene or gene fragments that compensate for or override defective gene expression such that cell phenotypes associated with defective gene expression are eliminated or changed to a different or desired cellular phenotype.

[0003] In an embodiment of the invention, the donor polynucleotide may include, but not be limited to, genes or gene fragments, encoding proteins or RNA transcripts to be expressed, regulatory elements, repair templates, and the like. According to the invention, the donor polynucleotides may comprise left end and right end sequence elements that function with transposition components that mediate insertion.

[0004] In certain cases, the donor polynucleotide manipulates a splicing site on the target polynucleotide. In some examples, the donor polynucleotide disrupts a splicing site. The disruption may be achieved by inserting the polynucleotide to a splicing site and/or introducing one or more mutations to the splicing site. In certain examples, the donor polynucleotide may restore a splicing site. For example, the polynucleotide may comprise a splicing site sequence.

[0005] The donor polynucleotide to be inserted may has a size from 10 base pair or nucleotides to 50 kb in length, e.g., from 50 to 40k, from 100 and 30 k, from 100 to 10000,

from 100 to 300, from 200 to 400, from 300 to 500, from 400 to 600, from 500 to 700, from 600 to 800, from 700 to 900, from 800 to 1000, from 900 to from 1100, from 1000 to 1200, from 1100 to 1300, from 1200 to 1400, from 1300 to 1500, from 1400 to 1600, from 1500 to 1700, from 600 to 1800, from 1700 to 1900, from 1800 to 2000 base pairs (bp) or nucleotides in length.

Functional Domain Modifications

[0309] The chimeric IscB systems described above may be further functionalized via their association with one or more functional domains. The chimeric IscB systems may be further modified such that they function a nickases, as detailed herein, or rendered catalytically inactive by mutating one or more residues in the catalytic site of the nuclease. These catalytically inactive or nickase forms of the chimeric IscB may then be paired with other functional domains to extend the functionality of these IscB chimeric systems. Example functional domains include nucleotide deaminase, reverse transcriptase, non-LTR retrotransposon (and protein encoded), polymerase, diversity generating element (and protein encoded). The functional domains may be covalently linked to the chimeric IscB or otherwise configured so as to be able to associate with the chimeric IscB in solution or in a cell.

Base Editors

[0310] The chimeric IscB systems disclosed herein may be further modified to include nucleotide deaminase (e.g., an adenosine deaminase or cytidine deaminase) associated (e.g., fused) with the chimeric IscB system, (e.g., IscB protein.). In certain examples, the nucleotide deaminase is a mutated form of an adenosine deaminase The mutated form of the adenosine deaminase may have both adenosine deaminase and cytidine deaminase activities.

[0311] In some examples, the present disclosure provides an engineered, non-naturally occurring composition comprising: the nucleic acid-guided nuclease that is catalytically inactive, a nucleotide deaminase associated with or otherwise capable of forming a complex with the IscB protein, and a single ωRNA molecule or single guide RNA molecule capable of forming a complex with the IscB protein and directing site-specific binding at a target sequence.

[0312] In one aspect, the present disclosure provides an engineered adenosine deaminase. The engineered adenosine deaminase may comprise one or more mutations herein. In one embodiment, the engineered adenosine deaminase has cytidine deaminase activity. In certain examples, the engineered adenosine deaminase has both cytidine deaminase activity and

adenosine deaminase. In some cases, the modifications by base editors herein may be used for targeting post-translational signaling or catalysis. In one embodiment, compositions herein comprise nucleotide sequence comprising encoding sequences for one or more components of a base editing system. A base-editing system may comprise a deaminase (e.g., an adenosine deaminase or cytidine deaminase) fused with a chimeric IscB system or a variant thereof. In some cases, the target polynucleotide is edited at one or more bases to introduce a G→A or C→T mutation.

[0313] In some cases, the adenosine deaminase is double-stranded RNA-specific adenosine deaminase (ADAR). Examples of ADARs include those described Yiannis A Savva et al., The ADAR protein family, Genome Biol. 2012; 13(12): 252, which is incorporated by reference in its entirety. In some examples, the ADAR may be hADAR1. In certain examples, the ADAR may be hADAR2. The sequence of hADAR2 may be that described under Accession No. AF525422.1.

[0314] In some cases, the deaminase may be a deaminase domain, e.g., a deaminase domain of ADAR ("ADAR-D"). In one example, the deaminase may be the deaminase domain of hADAR2 ("hADAR2-D), e.g., as described in Phelps KJ et al., Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. Nucleic Acids Res. 2015 Jan;43(2):1123-32, which is incorporated by reference herein in its entirety. In a particular example, the hADAR2-D has a sequence comprising amino acid 299-701 of hADAR2-D, e.g., amino acid 299-701 of the sequence under Accession No. AF525422.1.

[0315] In certain examples, the system comprises a mutated form of an adenosine deaminase fused with a dead chimeric IscB system (e.g., a IscB polypeptide nickase). The mutated form of the adenosine deaminase may have both adenosine deaminase and cytidine deaminase activities. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the

mutations: E488Q, V351G, S486A, T375S, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one

or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising one or more mutations of E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T, fused with a dead chimeric IscB system or IscB polypeptide nickase. In some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, and S661T, fused with a dead chimeric IscB system or IscB polypeptide nickase. In some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T, and S375N fused with a dead chimeric IscB system or IscB polypeptide nickase.

[0316] In one embodiment, the adenosine deaminase may be a tRNA-specific adenosine deaminase or a variant thereof. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: W23L, W23R, R26G, H36L, N37S, P48S, P48T, P48A, I49V, R51L, N72D, L84F, S97C, A106V, D108N, H123Y, G125A, A142N, S146C, D147Y, R152H, R152P, E155V, I156F, K157N, K161T, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one

embodiment, the adenosine deaminase may comprise one or more of the mutations: D108N based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, A142N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, A142N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V,

D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, A142N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, R152P, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, R152P, A142N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above.

[0317] In some examples, the base editing systems may comprise an intein-mediated trans-splicing system that enables *in vivo* delivery of a base editor, e.g., a split-intein cytidine base editors (CBE) or adenine base editor (ABE) engineered to trans-splice. Examples of such base editing systems include those described in Colin K.W. Lim et al., Treatment of a Mouse Model of ALS by *In Vivo* Base Editing, Mol Ther. 2020 Jan 14. pii: S1525-0016(20)30011-3. doi: 10.1016/j.ymthe.2020.01.005; and Jonathan M. Levy et al., Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses, Nature Biomedical Engineering volume 4, pages97–110(2020), which are incorporated by reference herein in their entireties.

[0318] Examples of base editing systems include those described in International Patent Publication Nos. WO 2019/071048 (e.g. paragraphs [0933]-[0938]), WO 2019/084063 (e.g., paragraphs [0173]-[0186], [0323]-[0475], [0893]-[1094]), WO 2019/126716 (e.g., paragraphs [0290]-[0425], [1077]-[1084]), WO 2019/126709 (e.g., paragraphs [0294]-[0453]), WO 2019/126762 (e.g., paragraphs [0309]-[0438]), WO 2019/126774 (e.g., paragraphs [0511]-[0670]), Cox DBT, et al., RNA editing with CRISPR-Cas13, Science. 2017 Nov 24;358(6366):1019-1027; Abudayyeh OO, et al., A cytosine deaminase for programmable single-base RNA editing, Science 26 Jul 2019: Vol. 365, Issue 6451, pp. 382-386; Gaudelli NM et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, Nature volume 551, pages 464–471 (23 November 2017); Komor AC, et al., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May 19;533(7603):420-4; Jordan L. Doman et al., Evaluation and minimization of Cas9-

independent off-target DNA editing by cytosine base editors, Nat Biotechnol (2020). doi.org/10.1038/s41587-020-0414-6; and Richter MF et al., Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity, Nat Biotechnol (2020). doi.org/10.1038/s41587-020-0453-z, Gaudelli et al., Directed evolution of adenine base editors with increased activity and therapeutic application, Nature Biotech. 38, 892-900 (2020), which are incorporated by reference herein in their entireties and can be used to adapt to the chimeric IscB system.

Prime Editors

[0319] In one embodiment, the present disclosure provides compositions and systems may comprise a chimeric IscB system or a catalytically inactive form, one or more ωRNA or guide molecules, and a reverse transcriptase. The systems may be used to insert a donor polynucleotide to a target polynucleotide. In some examples, the composition or system comprises a catalytically inactive chimeric IscB system, a reverse transcriptase associated with or otherwise capable of forming a complex with the chimeric IscB systems, and a ωRNA or guide molecule capable of forming a complex with the chimeric IscB systems and directing site-specific binding of the complex to a target sequence of a target polynucleotide, the ωRNA or guide molecule further comprising a donor sequence for insertion into the target polynucleotide.

[0320] In some cases, the catalytically inactive chimeric IscB systems may be a nickase, e.g., a DNA nickase. In some cases, the chimeric IscB system has one or more mutations. In some examples, the chimeric IscB system comprises mutations corresponding to the mutations in the RuvC or HNH nuclease.

[0321] The chimeric IscB systems may be associated with a reverse transcriptase. As used in this context "associated" means covalently linked, e.g. via a linker, or otherwise capable of complexing with the reverse transcriptase. A reverse transcriptase domain may be a reverse transcriptase or a fragment thereof.

[0322] In some examples, the compositions and systems may comprise the chimeric IscB system disclosed herein; a reverse transcriptase (RT) polypeptide connected to or otherwise capable of forming a complex with the chimeric IscB system; and a ωRNA or guide molecule capable of forming a complex with the chimeric IscB system and comprising: a ωRNA or guide sequence capable of directing site-specific binding of the chimeric IscB system complex to a target sequence of a target polynucleotide; a 3' binding site region capable of binding to a

cleaved upstream strand of the target polynucleotide; and a RT template sequence encoding an extended sequence, wherein the extended sequence comprises a variant region and a 3' homologous sequence capable of hybridization to the downstream cleaved strand of the target polynucleotide.

[0323] A reverse transcriptase domain may be a reverse transcriptase or a fragment thereof. A wide variety of reverse transcriptases (RT) may be used in alternative embodiments of the present invention, including prokaryotic and eukaryotic RT, provided that the RT functions within the host to generate a donor polynucleotide sequence from the RNA template. If desired, the nucleotide sequence of a native RT may be modified, for example using known codon optimization techniques, so that expression within the desired host is optimized. A reverse transcriptase (RT) is an enzyme used to generate complementary DNA (cDNA) from an RNA template, a process termed reverse transcription. Reverse transcriptases are used by retroviruses to replicate their genomes, by retrotransposon mobile genetic elements to proliferate within the host genome, by eukaryotic cells to extend the telomeres at the ends of their linear chromosomes, and by some non-retroviruses such as the hepatitis B virus, a member of the Hepadnaviridae, which are dsDNA-RT viruses. Retroviral RT has three sequential biochemical activities: RNA-dependent DNA polymerase activity, ribonuclease H, and DNA-dependent DNA polymerase activity. Collectively, these activities enable the enzyme to convert single-stranded RNA into double-stranded cDNA. In an embodiment, the RT domain of a reverse transcriptase is used in the present invention. The domain may include only the RNA-dependent DNA polymerase activity. In one embodiment, the RT domain is non-mutagenic, i.e., does not cause mutation in the donor polynucleotide (e.g., during the reverse transcriptase process). In example embodiments, the RT domain may be non-retron RT, e.g., a viral RT or a human endogenous RTs. In some examples, the RT domain may be retron RT or DGRs RT. In some examples, the RT may be less mutagenic than a counterpart wildtype RT. In one embodiment, the RT herein is not mutagenic. In one embodiment, the reverse transcriptase is Human immunodeficiency virus (HIV) RT, Avian myoblastosis virus (AMV) RT, Moloney murine leukemia virus (M-MLV) RT a group II intron RT, a group II intron-like RT, or a chimeric RT. In an embodiment, the RT comprises modified forms of these RTs, such as, engineered variants of Avian myoblastosis virus (AMV) RT, Moloney murine leukemia virus (M-MLV) RT, or Human immunodeficiency virus (HIV) RT (see, e.g., Anzalone, et al.,

Search-and-replace genome editing without double-strand breaks or donor DNA, Nature. 2019 Dec;576(7785):149-157).

[0324] The reverse transcriptase may be fused to the C-terminus of a chimeric IscB system. Alternatively or additionally, the reverse transcriptase may be fused to the N-terminus of a chimeric IscB system. The fusion may be via a linker and/or an adaptor protein. In some examples, the reverse transcriptase may be an M-MLV reverse transcriptase or variant thereof. The M-MLV reverse transcriptase variant may comprise one or more mutations. For the examples, the M-MLV reverse transcriptase may comprise D200N, L603W, and T330P. In another example, the M-MLV reverse transcriptase may comprise D200N, L603W, T330P, T306K, and W313F. In a particular example, the fusion of chimeric IscB systems and reverse transcriptase is chimeric IscB system (with a mutation corresponding to H840A of SpCas9) fused with M-MLV reverse transcriptase (D200N+L603W+T330P+T306K+W313F).

[0325] In one embodiment, the chimeric IscB systems herein may target DNA using a ωRNA or guide RNA containing a binding sequence that hybridizes to the target sequence on the DNA. The ωRNA or guide RNA may further comprise an editing sequence that contains new genetic information that replaces target DNA nucleotides. The small sizes of the chimeric IscB systems herein may allow easier packaging and delivery of the prime editing system, e.g., with a viral vector, e.g., AAV or lentiviral vector.

[0326] A single-strand break (a nick) may be generated on the target DNA by the chimeric IscB systems at the target site to expose a 3'-hydroxyl group, thus priming the reverse transcription of an edit-encoding extension on the ωRNA or guide directly into the target site. These steps may result in a branched intermediate with two redundant single-stranded DNA flaps: a 5' flap that contains the unedited DNA sequence, and a 3' flap that contains the edited sequence copied from the ωRNA. The 5' flaps may be removed by a structure-specific endonuclease, e.g., FEN122, which excises 5' flaps generated during lagging-strand DNA synthesis and long-patch base excision repair. The non-edited DNA strand may be nicked to induce bias DNA repair to preferentially replace the non-edited strand. Examples of prime editing systems and methods include those described in Anzalone AV *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA, Nature. 2019 Oct 21. doi: 10.1038/s41586-019-1711-4, which is incorporated by reference herein in its entirety.

[0327] The chimeric IscB system (e.g., the nickase form) may be used to prime-edit a single nucleotide on a target DNA. Alternatively or additionally, the chimeric IscB systems may be

used to prime-edit at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, 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, at least 25, at least 26, at least 27, at least 28, at least 29, 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 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 1000 nucleotides on a target DNA.

[0328] Examples of prime editing systems and methods that may be adapted for use with the chimeric IscBs described herein include those described in Anzalone x., "Search-and-replace genome editing without double-strand breaks or donor DNA", Nature. 576, 149-157 (2019); Chen *et al.* "Enhanced prime editing systems by manipulating cellular determinants of editing outcomes" Cell 184(22):5635-5652.e29 (2021); WO 2020/191233; WO 2020/191234; WO 2020/191239; WO 2020/191241; WO 2020/191242; WO 2020/191243; WO 2020/191245; WO 2020/191246; WO 2020/191248; WO 2020/191249; WO 2021/082328 4, which is incorporated by reference herein in its entirety. In such cases, the system comprises a chimeric IscB system with nickase activity, a reverse transcriptase domain, and a DNA polymerase, and a ωRNA or guide molecule comprising a binding sequence capable of hybridizing to the target polynucleotide and an editing sequence. The generated region may be further extended on a DNA template as described herein. The latter may allow generation of a target-independent sequence, compatible with a generic donor sequence.

[0329] The chimeric IscB system is capable of generating a first cleavage of in the target sequence and a second cleavage outside the target sequence on the target polynucleotide. In some variations, a second chimeric IscB system-mediated cleavage in vicinity to the target site may be made, which may enable more efficient invasion of the extended DNA.

[0330] In one example embodiment, multiple chimeric IscB prime editing systems may be used in combination to facilitate larger insertions and deletions. In one embodiment, the compositions and systems of the chimeric IscB system herein comprise: a reverse transcriptase (RT) polypeptide connected to or otherwise capable of forming a complex with the chimeric IscB system; a first ωRNA or guide molecule capable of forming a first chimeric IscB system-Reverse transcriptase complex with the chimeric IscB system and comprising: a ωRNA or guide sequence capable of directing site-specific binding of the first chimeric IscB system-Reverse transcriptase complex to a first target sequence of a target polynucleotide; a first binding site region capable of binding to a cleaved or nicked strand of the target polynucleotide;

and a RT template sequence encoding a first extended sequence; a second ωRNA or guide molecule capable of forming a second chimeric IscB system-Reverse transcriptase complex with the chimeric IscB system and comprising: a ωRNA or guide sequence capable of directing site specific binding of the second chimeric IscB system-Reverse transcriptase complex to a second target sequence of the target polynucleotide; a second binding site region capable of binding to a cleaved or nicked strand of the target polynucleotide; and a RT template sequence encoding a second extended sequence.

[0331] In some cases, the compositions and systems may further comprise: a donor template; a third ωRNA or guide sequence capable of forming a chimeric IscB system-Reverse transcriptase complex-ωRNA or guide with the chimeric IscB system and comprising: a ωRNA or guide sequence capable of directing site-specific binding to a target sequence on the donor template; a third binding region capable of binding to a cleaved or nicked strand of the donor template; and a RT template encoding a third extended region complementary to the first extended region generated on the target polynucleotide: and a fourth ωRNA or guide sequence capable of forming a chimeric IscB system-Reverse transcriptase complex with the IscB polypeptide or chimeric IscB system and comprising: a ωRNA or guide sequence capable of directing site-specific binding to a second target sequence on the donor template; a fourth binding region capable of binding to a cleaved or nicked strand of the donor template; and a RT template encoding a fourth extended region complementary to the second extended region generated on the target polynucleotide.

[0332] The use of two chimeric IscB systems (which may also be referred to as double-flap prime editing or twinPE) may be used to insert, delete, or replace larger sequences. Examples of CRISPR-Cas based prime editing systems that may be adapted for use with the chimeric IscBs described herein are disclosed in WO 2021/138469; Anzalone *et al.* "Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing" Nature Biotechnology 40(5):731-740 (2021); WO 20221/226558; WO 2021/ 226558, which are incorporated herein in their entirety by reference.

[0333] In another embodiment, the chimerc IscB prime editing compositions and systems may further comprise a site-specific recombinase. The recombinase is connected to or otherwise capable of forming a complex with the chimeric IscB prime editing system. In an embodiment, the complex is capable of inserting a recombination site in the DNA loci of interest by extension of RT templates that encode for the recombination site on the 3' extension

of the ωRNA or guide sequences by the reverse transcriptase. In an embodiment, a donor template comprising a compatible recombination site is provided that can recombine unidirectionally with the inserted recombination site when a recombinase specific for the recombination site is also provided. In an embodiment, the donor template is a plasmid comprising the complementary recombination site and any sequence for insertion at the DNA loci of interest. In an embodiment, the recombinase is connected to or capable of forming a complex with the chimeric IscB systems, such that all of the enzymatic proteins are brought into contact at the loci of interest. In an embodiment, the recombinase is codon optimized for eukaryotic cells (described further herein). In an embodiment, the recombinase includes a NLS (described further herein). In an embodiment, the recombinase is provided as a separate protein. The separate recombinase may form a dimer and bind to the donor template recombination site. The recombinase may be targeted to the loci of interest as a result of the insertion of the compatible recombination site that is also recognized by the recombinase. Thus, the recombinase may recognize the recombination site inserted at the DNA loci of interest and the recombination site on the donor and be targeted to the DNA loci of interest without any additional modifications to the recombinase.

[0334] In an embodiment, a second IscB complex connected to a recombinase is targeted to the DNA loci of interest. In an embodiment, the second TnpB complex comprises a dead IscB protein (dIscB, described further herein), such that the recombinase is targeted to the DNA loci of interest, but the target sequence is not further cleaved. In an embodiment, the dIscB targets a sequence generated only after the insertion of the recombination site. In an embodiment, the recombinase recognizes and binds to the donor template recombination site and the inserted recombination site. In an embodiment, the recombinase forms a dimer with a recombinase provided as a separate protein.

[0335] As used herein, the term "Recombinase" refers to an enzyme that catalyzes recombination between two or more recombination sites (e.g., an acceptor and donor site). Recombinases useful in the present invention catalyze recombination at specific recombination sites which are specific polynucleotide sequences that are recognized by a particular recombinase. "Uni-directional recombinases" or "integrases" refer to recombinase enzymes whose recognition sites are destroyed after the recombination has taken place. The term "integrase" refers to a type of recombinase. In other words, the sequence recognized by the recombinase is changed into one that is not recognized by the recombinase upon

recombination. As a result, once a sequence is subjected to recombination by the uni-directional recombinase, the continued presence of the recombinase cannot reverse the previous recombination event.

[0336] "Recombination sites" are specific polynucleotide sequences that are recognized by the recombinase enzymes described herein. Typically, two different sites are involved (in regards to recombination termed "complementary sites"), one present in the target nucleic acid (e.g., a chromosome or episome of a eukaryote) and another on the nucleic acid that is to be integrated at the target recombination site. The terms "attB" and "attP," which refer to attachment (or recombination) sites originally from a bacterial target (attachment site of bacteria) and a phage donor (attachment site of phage), respectively, are used herein although recombination sites for particular enzymes may have different names. The two attachment sites can share as little sequence identity as a few base pairs. The recombination sites typically include left and right arms separated by a core or spacer region. Thus, an attB recombination site consists of BOB′, where B and B′ are the left and right arms, respectively, and O is the core region. Similarly, attP is POP′, where P and P′ are the arms and O is again the core region. Upon recombination between the attB and attP sites, and concomitant integration of a nucleic acid at the target, the recombination sites that flank the integrated DNA are referred to as "attL" and "aatR." The attL and attR sites, using the terminology above, thus consist of BOP′ and POB′, respectively. In some representations herein, the "O" is omitted and attB and attP, for example, are designated as BB′ and PP′, respectively.

[0337] Example CRISPR-Cas prime editing/recombinase compositions and systems that may be adapted for use with the chimeric IscBs disclosed herein are described in WO 2021/138469, Anzalone 2021; WO 2021/226558; Yarnall *et al.* "Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases" 41, 500-512 (2023); and WO 2022/087235.

Transposition Systems

[0338] In another aspect, the chimeric IscB systems disclosed herein may further comprise a transposase and optionally a donor template. The chimeric IscB may be catalytically inactive. The chimeric IscB maybe further linked to, or otherwise capable of associating with, a transposase (or one or more components thereof). The chimeric IscB may then direct the transposase to a desired transposition site where the transposase facilitate insertion of a donor polynucleotide e.g. from the provided donor template into the desired transposition site.

[0339] The term "transposase" as used herein refers to an enzyme, which is a component of a functional nucleic acid-protein complex capable of transposition and which mediates transposition. The transposase may comprise a single protein or comprise multiple protein sub-units. A transposase may be an enzyme capable of forming a functional complex with a transposon end or transposon end sequences. The term "transposase" may also refer in certain embodiments to integrases. The expression "transposition reaction" used herein refers to a reaction wherein a transposase inserts a donor polynucleotide sequence in or adjacent to an insertion site on a target polynucleotide. The insertion site may contain a sequence or secondary structure recognized by the transposase and/or an insertion motif sequence where the transposase cuts or creates staggered breaks in the target polynucleotide into which the donor polynucleotide sequence may be inserted. Exemplary components in a transposition reaction include a transposon, comprising the donor polynucleotide sequence to be inserted, and a transposase or an integrase enzyme. The term "transposon end sequence" as used herein refers to the nucleotide sequences at the distal ends of a transposon. The transposon end sequences may be responsible for identifying the donor polynucleotide for transposition. The transposon end sequences may be the DNA sequences the transpose enzyme uses in order to form transpososome complex and to perform a transposition reaction.

[0340] Embodiments disclosed herein provide an engineered or non-natural guided excision-transposition system. The engineered or non-natural guided excision-transposition system may comprise one or more components of a ωRNA-chimeric IscB system and one or more components of a Class II transposon. The components of the ωRNA-IscB or guide-chimeric IscB system can direct the Class II transposon component(s) to retrotransposon to a target nucleic acid sequence and direct its transposition into a recipient polynucleotide.

[0341] For example, the engineered or non-natural guided excision-transposition systems that can include (a) a first chimeric IscB system; (b) a first Class II transposon polypeptide coupled to or otherwise capable of complexing with the first chimeric IscB system; (c) a first guide molecule capable of forming a first ωRNA-IscB or guide-chimeric IscB system with the first chimeric IscB system and directing site-specific binding to a first target sequence of a first target polynucleotide; (d) a second chimeric IscB system; (e) a second Class II transposon polypeptide coupled to or otherwise capable of complexing with the second chimeric IscB system; (f) a second guide molecule capable of forming a second ωRNA-IscB complex with the first IscB protein and directing site-specific binding to a second target sequence of the first

target polynucleotide; and (g) a Class II transposon polynucleotide comprising the first target polynucleotide and is capable of forming a complex with the first and second chimeric IscB system, the first and second guide molecules, and the first and second Class II transposon polypeptides.

[0342] In one embodiment, the engineered or non-natural guided excision-transposition system can include (h) a third guide molecule capable of complexing with the first chimeric IscB system and directing site-specific binding to a first target sequence of a second target polynucleotide, wherein the third guide molecule is optionally coupled to the first chimeric IscB system; (i) optionally, a first ωRNA or guide molecule polynucleotide that encodes the third ωRNA or guide molecule; (j) a fourth ωRNA or guide molecule capable of complexing with the second chimeric IscB system and directing site-specific binding to a second target sequence of the second target polynucleotide, wherein the fourth guide molecule is optionally coupled to the second chimeric IscB system; and (k) optionally, a second ωRNA or guide molecule polynucleotide that encodes the fourth ωRNA or guide molecule.

[0343] In one embodiment, the first and the second Class II transposon polypeptides are capable of excising the first target polynucleotide from the Class II transposon polynucleotide. In one embodiment, the first and the second Class II transposon polypeptides are capable of transposing the first target polynucleotide in the second target polynucleotide. In one embodiment, the first target polynucleotide does not include one or more Class II transposon long terminal repeats.

[0344] The engineered or non-natural guided excision-transposition systems described herein can be based on a Class II transposon or Class II transposon system. The engineered or non-natural guided excision-transposition system may include a first target polynucleotide, also referred to as a donor polynucleotide or transposon and a second target polynucleotide, which is also referred to herein as a recipient polynucleotide. As used herein, "transposon" (also referred to as transposable element) refers to a polynucleotide sequence that is capable of moving form location in a genome to another. There are several classes of transposons. Transposons include retrotransposons (Class I transposons) and DNA transposons (Class II transposons). In some cases, retrotransposons require the transcription of the polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide. DNA transposons are those that do not require reverse transcription of the

polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide.

[0345] Any suitable transposon system can be used. Suitable transposon and systems thereof can include, but are not limited, to Sleeping Beauty transposon system (Tc1/mariner superfamily) (see e.g. Ivics et al. 1997. Cell. 91(4): 501-510), piggyBac (piggyBac superfamily) (see e.g. Li et al. 2013 110(25): E2279-E2287 and Yusa et al. 2011. PNAS. 108(4): 1531-1536), Tol2 (superfamily hAT), Frog Prince (Tc1/mariner superfamily) (see e.g., Miskey et al. 2003 Nucleic Acid Res. 31(23):6873-6881) and variants thereof.

[0346] In one embodiment, the first and/or second Class II transposon polypeptide is a DD[E/D] transposon or transposon polypeptide. In one embodiment, the first and/or the second Class II transposon polynucleotide is a Tc1/mariner, PiggyBac, Frog Prince, Tn3, Tn5, hAT, CACTA, P, Mutator, PIF/Harbinger, Transib, or a Merlin/IS1016 transposon polynucleotide. In one embodiment, the first and/or second Class II transposon polypeptide is a Tc1/mariner, PiggyBac, Frog Prince, Tn3, Tn5, hAT, CACTA, P, Mutator, PIF/Harbinger, Transib, or a Merlin/IS1016 transposon polypeptide.

[0347] Suitable Class II transposon systems and components that can be utilized can also be and are not limited to those described in e.g., and without limitation, Han et al., 2013. BMC Genomics. 14:71, doi: 10.1186/1471-2164-14-71, Lopez and Garcia-Perez. 2010. Curr. Genomics. 11(2):115-128; Wessler. 2006. PNAS. 103(47): 176000-17601; Gao et al., 2017. Marine Genomics. 34:67-77; Bradic et al. 2014. Mobile DNA. 5(12) doi:10.1186/1759-8753-5-12; Li et al., 2013. PNAS. 110(25)E2279-E2287; Kebriaei et al. 2017. Trends in Genetics. 33(11): 852-870); Miskey et al. 2003. Nucleic Acid res. 31(23):6873-6881; Nicolas et al. 2015. Microbiol Spectr. 3(4) doi: 10.1128/microbiolspec.MDNA3-0060-2014); W.S. Reznikoff. 1993. Annu Rev. Microbiol. 47:945-963; Rubin et al. 2001. Genetics. 158(3): 949-957; Wicker et al. 2003. Plant Physiol. 132(1): 52-63; Majumdar and Rio. 2015. Microbiol. Spectr. 3(2) doi: 10.1128/microbiolspec.MDNA3-0004-2014; D. Lisch. 2002. Trends in Plant Sci. 7(11): 498-504; Sinzelle et al. 2007. PNAS. 105(12): 4715-4720; Han et al. 2014; Genome Biol. Evol. 6(7):1748-1757; Grzebelus et al. 2006; Mol. Genet. Genomics. 275(5):450-459; Zhang et al. 2004. Genetics. 166(2):971-986; Chen and Li. 2008. Gene. 408(1-2):51-63; and C. Feschotte. 2004. Mol. Biol. Evol. 21(9):1769-1780.

T7 transposases

[0348] In some embodiments, the system comprises one or more Tn7 or Tn7-like transposases. In a particular embodiment, the Tn7-like transposase may be a Tn5053 transposase. For example, the Tn5053 transposases include those described in Minakhina S et al., Tn5053 family transposons are res site hunters sensing plasmidal res sites occupied by cognate resolvases. Mol Microbiol. 1999 Sep;33(5):1059-68; and Fig. 4 and related texts in Partridge SR et al., Mobile Genetic Elements Associated with Antimicrobial Resistance, Clin Microbiol Rev. 2018 Aug 1;31(4), both of which are incorporated by reference herein in their entirety. In some cases, the one or more Tn5053 transposases may comprise one or more of TniA, TniB, and TniQ. TniA is also known as TnsB. TniB is also known as TnsC. TniQ is also known as TnsD. Accordingly, in certain embodiments these Tn5053 transposase subunits may be referred to as TnsB, TnsC, and TnsD, respectively. In certain cases, the one or more transposases may comprise TnsB, TnsC, and TnsD. In one example, a chimeric IscB system comprises TniA, TniB, TniQ, Cas12k, tracrRNA, and guide RNA(s). In another example, a chimeric IscB system comprises TnsB, TnsC, TnsD, Cas12k, tracrRNA, and guide RNA(s).

[0349] In some examples, the one or more transposases may comprise: (a) TnsA, TnsB, TnsC, and TniQ, (b) TnsA, TnsB, and TnsC, (c) TnsB and TnsC, (d) TnsB, TnsC, and TniQ, (e) TnsA, TnsB, and TniQ, (f) TnsE, or (g) any combination thereof. In some cases, the TnsE does not bind to DNA. In some cases, transposase protein may comprise one or more transposases, e.g., one or more transposase subunits of a Tn7 transposase or Tn7-like transposase, e.g., one or more of TnsA, TnsB, TnsC, and TniQ. In some examples, the one or more transposases comprise TnsB, TnsC, and TniQ.

[0350] In some embodiments, three transposon-encoded proteins form the core transposition machinery of Tn7: a heteromeric transposase (TnsA and TnsB) and a regulator protein (TnsC). In addition to the core TnsABC transposition proteins, Tn7 elements encode dedicated target site-selection proteins, TnsD and TnsE. In conjunction with TnsABC, the sequence-specific DNA-binding protein TnsD directs transposition into a conserved site referred to as the "Tn7 attachment site," attTn7. TnsD is a member of a large family of proteins that also includes TniQ, a protein found in other types of bacterial transposons. TniQ has been shown to target transposition into resolution sites of plasmids. As used herein, a TniQ transposase may be a TnsD transposase. Examples of Tn7 or Tn7-like transposases include TnsA, TnsB, TnsC, TniQ, TnsD, and TnsE. In some embodiments, the system comprises TnsA,

TnsB, TnsC, and/or TniQ. Two or more of the components in the system may be comprised in a single protein (e.g., fusion protein). For example, TnsA and TnsB may be comprised in a single protein.

[0351] As used herein, a right end sequence element or a left end sequence element are made in reference to an example Tn7 transposon. The general structure of the left end (LE) and right end (RE) sequence elements of canonical Tn7 is established. Tn7 ends comprise a series of 22-bp TnsB-binding sites. Flanking the most distal TnsB-binding sites is an 8-bp terminal sequence ending with 5′-TGT-3′/3′-ACA-5′. The right end of Tn7 contains four overlapping TnsB-binding sites in the ~90-bp right end element. The left end contains three TnsB-binding sites dispersed in the ~150-bp left end of the element. The number and distribution of TnsB-binding sites can vary among Tn7-like elements. End sequences of Tn7-related elements can be determined by identifying the directly repeated 5-bp target site duplication, the terminal 8-bp sequence, and 22-bp TnsB-binding sites (Peters JE et al., 2017). Example Tn7 elements, including right end sequence element and left end sequence element include those described in Parks AR, Plasmid, 2009 Jan; 61(1):1-14.

Tn5 Transposases

[0352] In certain embodiments, the one or more transposases are one or more Tn5 transposases. In some examples, the transposases may comprise TnpA. The transposase may be a Y1 transposase of the IS200/IS605 family, encoded by the insertion sequence (IS) IS608 from Helicobacter pylori, e.g., TnpAIS608. Examples of the transposases include those described in Barabas, O., Ronning, D.R., Guynet, C., Hickman, A.B., TonHoang, B., Chandler, M. and Dyda, F. (2008) Mechanism of IS200/ IS605 family DNA transposases: activation and transposon-directed target site selection. Cell, 132, 208–220. In certain example embodiments, the transposase is a single stranded DNA transposase. In certain example embodiments, the single stranded DNA transposase is TnpA or a functional fragment thereof. The chimeric IscB - transposase systems may be coded on the same strand or be part of a larger operon. In certain embodiments, the chimeric IscB may confer target specificity, allowing the TnpA to move a polynucleotide cargo from other target sites in a sequence specific matter. In certain example embodiments, the transposase are derived from *Flavobactreium granuli* strain DSM-19729, *Salinivirga cyanobacteriivorans* strain L21-Spi-D4, *Flavobactrium aciduliphilum* strain DSM 25663, *Flavobacterium glacii* strain DSM 19728, *Niabella soli* DSM 19437, *Salnivirga*

cyanobactriivorans strain L21-Spi-D4, *Alkaliflexus imshenetskii* DSM 150055 strain Z-7010, or *Alkalitala saponilacus*.

[0353] In certain embodiments, the transposase is a single-stranded DNA transposase. The single stranded DNA transposase may be TnpA, a functional fragment thereof, or a variant thereof. In certain embodiments, the transposase is a Himar1 transposase, a fragment thereof, or a variant thereof. In one example, the system comprises a chimeric IscB associated with Himar1.

[0354] In certain embodiments, the transposases may be one or more *Vibrio cholerae* Tn6677 transposases. In one example, the system may comprise components of variant Type I-F CRISPR–Cas system or polynucleotide(s) encoding thereof. The transposon may include a terminal operon comprising the tnsA, tnsB, and tnsC genes. The transposon may further comprise a tniQ gene. The tniQ gene may be encoded within the cas rather than tns operon. In certain embodiments, the TnsE may be absent in the transposon.

Mu transposases

[0355] The transposases may be one of the Mu family transposon systems, e.g., transposon of bacteriophage Mu, a bacterial class III transposon of Escherichia coli. In some cases, this transposon exhibits high transposition frequency. The Mu bacteriophage with its approximately 37 kb genome is relatively large compared to other transposons. The Mu transposon may have left end and right end transposase (e.g., MuA) recognition sequences (designated "L" and "R", respectively) that flank the Mu transposable cassette, the region of the transposon that is ultimately integrated into the target site. In some examples, these ends are not inverted repeat sequences. The Mu transposable cassette, when necessary, may include a transpositional enhancer sequence (also referred to herein as the internal activating sequence, or "IAS") located approximately 950 base pairs inward from the left end recognition sequence.

[0356] In some examples, a Mu transposon may have a 22 bp symmetrical consensus sequence, located near both ends, for recognition by a Mu transposase (MuA). Random transposition of a Mu transposon into a target gene occur through (1) binding of transposase (e.g., MuA) monomers to the Mu transposon recognition sites to form transposome assemblies, (2) tetramerization of the bound transposase (e.g., MuA) monomers to bridge the ends of the Mu transposon and engage the Mu transposon cleavage sites, (3) subsequent self-cleavage of the Mu transposon at the cleavage sites, and (4) accurate occurrence of a 5 bp staggered cut in a host DNA sequence into which the Mu transposon is subsequently incorporated.

[0357] The transposases may be Mu transposase family. Examples of transposases in the Mu family includes MuA, MuB, and MuC.

[0358] In some examples, MuA may be a about 75-kDa multidomain protein (about 663 amino acids) and can be divided into structurally and functionally defined major domains (I, II, III) and subdomains (Iα, Iβ, Iγ; IIα, IIβ; IIIα, IIIβ). The N-terminal subdomain Iα promotes transpososome assembly via an initial binding to a specific transpositional enhancer sequence. The specific DNA binding to transposon ends, crucial for the transpososome assembly, is mediated through amino acid residues located in subdomains Iβ and Iγ. Subdomain IIα contains the critical DDE-motif of acidic residues (D269, D336 and E392), which is involved in the metal ion coordination during the catalysis. Subdomains IIβ and IIIα participate in nonspecific DNA binding, and they appear important during structural transitions. Subdomain IIIα also displays a cryptic endonuclease activity, which is required for the removal of the attached host DNA following the integration of infecting Mu. The C-terminal subdomain IIIβ is responsible for the interaction with the phage-encoded MuB protein, important in targeting transposition into distal target sites. This subdomain is also important in interacting with the host-encoded C1pX protein, a factor which remodels the transpososome for disassemble.

[0359] In some examples, MuA may catalyze the steps of transposition: (i) initial cleavages at the transposon-host boundaries (donor cleavage) and (ii) covalent integration of the transposon into the target DNA (strand transfer). These steps may proceed via sequential structural transitions within a nucleoprotein complex, a transpososome, the core of which contains four MuA molecules and two synapsed transposon ends. In vivo, the critical MuA-catalyzed reaction steps may also involve the phage-encoded MuB targeting protein, host-encoded DNA architectural proteins (HU and IHF), certain DNA cofactors (MuA binding sites and transpositional enhancer sequence), as well as stringent DNA topology. The reaction steps mimicking Mu transposition into external target DNA can be reconstituted in vitro using MuA transposase, 50 bp Mu R-end DNA segments, and target DNA as the only macromolecular components.

[0360] In some examples, MuA and variants include those disclosed by EBI accession No. UNIPROT:Q58ZD8 which has 36% identity to wild type MuA protein; Naigamwalla et al., 1998, (Journal of Molecular Biology 282:265-274) (mutations in domain IIIa of the Mu transposase protein); Rasila et al., 2012, (Plos One, 7(5):E37922) (functional mapping of MuA

transposase family protein structures with scanning mutagenesis); WO 2010/099296 (hyperactive piggyback transposases).

[0361] In some examples, MuB may be an ATP-dependent DNA binding protein, which is required for efficient transposition in vivo. Bacteriophage Mu transposition may be influenced by the ATP-utilizing protein MuB. In vitro, the MuA transposase may direct insertions into targets that are bound by MuB. In some cases, there is no particular sequence specificity to MuB binding. However, its distribution on DNA may not be random: MuB binding to target molecules that already contain Mu sequences is specifically destabilized through an ATP-dependent mechanism (19). In some examples, MuB also stimulates the DNA-breakage and DNA-joining activities of MuA (Adzuma and Mizuuchi (1988) Cell 53:257-266; Baker et al. (1991) Cell 65:1003-1013; Maxwell et al. (1987) Proc. Natl. Acad. Sci. USA 84:699-703; Surette and Chaconas (1991) J. Biol Chem. 266:17306-17313; Surette et al. (1991) J. Biol. Chem. 266:3118-3124; and Wu and Chaconas (1992) J. Biol. Chem. 267:9552-9558; and Wu and Chaconas, (1994) J. Biol. Chem. 269:28829-28833).

IscB Retrotransposon Systems

[0362] The systems and compositions herein may comprise a chimeric IscB system, one or more ωRNAs or guide RNAs, and one or more components of a retrotransposon, e.g., a non-LTR retrotransposon. The one or more components of a retrotransposon include a retrotransposon protein and retrotransposon RNA. The systems and compositions may be used to insert a donor polynucleotide to a target polynucleotide. The systems and compositions may further comprise a donor polynucleotide.

[0363] In some examples, the present disclosure provides an engineered, non-naturally occurring composition comprising: a chimeric IscB system, a non-LTR retrotransposon protein associated with or otherwise capable of forming a complex with the chimeric IscB systems; a single ωRNA or guide capable of forming a complex with the chimeric IscB systems and directing site-specific binding to a target sequence of a target polynucleotide. The composition may further comprise a donor construct comprising a donor polynucleotide for insertion to the target polynucleotide and located between two binding elements capable of forming a complex with the non-LTR retrotransposon protein. In some cases, the chimeric IscB system is engineered to have nickase activity.

[0364] In some examples, the chimeric IscB system is fused to the N-terminus of the non-LTR retrotransposon protein. In some examples, the chimeric IscB system is fused to the C-terminus of the non-LTR retrotransposon protein.

[0365] The guides may direct the fusion protein to a target sequence 5' of the targeted insertion site, and wherein the chimeric IscB system generates a double-strand break at the targeted insertion site. The guides may direct the fusion protein to a target sequence 3' of the targeted insertion site, and wherein the chimeric IscB system generates a double-strand break at the targeted insertion site.

[0366] The donor polynucleotide may further comprise a polymerase processing element to facilitate 3' end processing of the donor polynucleotide sequence. The polymerase may be a DNA polymerase, e.g., DNA polymerase I. In some examples, the polymerase may be an RNA polymerase.

[0367] In some examples, the donor polynucleotide further comprises a homology region to the target sequence on the 5' end of the donor construct, the 3' end of the donor construct, or both. In some examples, the homology region is from 1 to 50, from 5 to 30, from 8 to 25, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 base pairs in length.

[0368] Native or wild-type non-LTR retrotransposons encode the protein machinery necessary for their self-mobilization. The non-LTR retrotransposon element comprises a DNA element integrated into a host genome. This DNA element may encode one or two open reading frames (ORFs). For example, the R2 element of Bombyx mori encodes a single ORF containing reverse transcriptase (RT) activity and a restriction enzyme-like (REL) domain. L1 elements encode two ORFs, ORF1 and ORF2. ORF1 contains a leucine zipper domain involved in protein-protein interactions and a C-terminal nucleic acid binding domain. ORF2 has a N-terminal apurinic/apyrimidinic endonuclease (APE), a central RT domain, and a C-terminal cysteine histidine rich domain. An example replicative cycle of a non-LTR retrotransposon may comprise transcription of the full-length retrotransposon element to generate an mRNA active element (retrotransposon RNA). The active element mRNA is translated to generate the encoded retrotransposon proteins or polypeptides. A ribonucleoprotein complex comprising the active element and retrotransposon protein, or polypeptide, is formed and this RNP facilitates integration of the active element into the genome. The RNA-transposase complex

nicks the genome. The 3' end of the nicked DNA serves as a primer to allow the reverse transcription of the transposon RNA into cDNA. Fourth, the transposase proteins integrate the cDNA into the genome.

[0369] Elements of these systems may be engineered to work within the context of the invention. For example, a non-LTR retrotransposon polypeptide may be fused to a site-specific nuclease. The binding elements that allow a non-LTR retrotransposon polypeptide to bind to the native retrotransposon DNA element, may be engineered into a donor construct to facilitate entry of a donor polynucleotide sequence into a target polypeptide.

[0370] In the present invention the protein component of the non-LTR retrotransposon may be connected to or otherwise engineered to form a complex with a site-specific nuclease. The retrotransposon RNA may be engineered to encode a donor polynucleotide sequence. Thus, in certain example embodiments, the chimeric IscB system, via formation of a chimeric IscB system complex with a guide sequence, directs the retrotransposon complex (e.g., the retrotransposon polypeptide(s) and retrotransposon RNA to a target sequence in a target polynucleotide, where the retrotransposon RNP complex facilitates integration of the donor polynucleotide sequence into the target polynucleotide. Accordingly, the one or more non-LTR retrotransposon components may comprise retrotransposon polypeptides, or function domains thereof, that facilitate binding of the retrotransposon RNA, reverse transcription of the retrotransposon RNA into cDNA, and/or integration of the donor polynucleotide into the target polynucleotide, as well as retrotransposon RNA elements modified to encode the donor polynucleotide sequence.

[0371] Examples of non-LTR retrotransposons include CRE, R2, R4, L1, RTE, Tad, R1, LOA, I, Jockey, CR1. In one example, the non-LTR retrotransposon is R2. In another example, the non-LTR retrotransposon is L1. Examples of non-LTR retrotransposons may include those described in Christensen SM et al., RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site, Proc Natl Acad Sci U S A. 2006 Nov 21;103(47):17602-7; Eickbush TH et al, Integration, Regulation, and Long-Term Stability of R2 Retrotransposons, Microbiol Spectr. 2015 Apr;3(2):MDNA3-0011-2014. doi: 10.1128/microbiolspec.MDNA3-0011-2014; Han JS, Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions, Mob DNA. 2010 May 12;1(1):15. doi: 10.1186/1759-8753-1-15; Malik HS et al., The age and evolution

of non-LTR retrotransposable elements, Mol Biol Evol. 1999 Jun;16(6):793-805, which are incorporated by reference herein in their entireties.

[0372] Examples of the non-LTR retrotransposon polypeptides also include R2 from *Clonorchis sinensis, or Zonotrichia albicollis*.

[0373] A non-LTR retrotransposon may comprise multiple retrotransposon polypeptides or polynucleotides encoding same. In one embodiment, the retrotransposon polypeptides may form a complex. For example, a non-LTR retrotransposon is a dimer, e.g., comprising two retrotransposon polypeptides forming a dimer. The dimer subunits may be connected or form a tandem fusion. A chimeric IscB system may be associate with (e.g., connected to) one or more subunits of such complex. In some examples, the non-LTR retrotransposon is a dimer of two retrotransposon polypeptides; one of the retrotransposon polypeptides comprises nuclease or nickase activity and is connected with a chimeric IscB system.

[0374] The retrotransposon polypeptides may comprise one or more modifications to, for example, enhance specificity or efficiency of donor polynucleotide recognition, target-primed template recognition (TPTR). The retrotransposon polypeptides may also comprise one or more truncations or excisions to remove domains or regions of wild-type protein to arrive at a minimal polypeptide that retain donor polynucleotide recognition and TPTR. In some example embodiments, the native endonuclease activity may be mutated to eliminate endonuclease activity.

[0375] In certain example embodiments, the modifications or truncations of the non-LTR retrotransposon peptide may be in a zinc finger region, a Myb region, a basic region, a reverse transcriptase domain, a cysteine-histidine rich motif, or an endonuclease domain.

[0376] A non-LTR retrotransposon may comprise polynucleotide encoding one or more retrotransposon RNA molecules. The polynucleotide may comprise one or more regulatory elements. The regulatory elements may be promoters. The regulatory elements and promoters on the polynucleotides include those described throughout this application. For example, the polynucleotide may comprise a pol2 promoter, a pol3 promoter, or a T7 promoter.

[0377] In some cases, the polynucleotide encodes a retrotransposon RNA with at least a portion of its sequence complementary to a target sequence. For example, the 3' end of the retrotransposon RNA may be complementary to a target sequence. The RNA may be complementary to a portion of a nicked target sequence. In one embodiment, a retrotransposon

RNA may comprise one or more donor polynucleotides. In certain cases, a retrotransposon RNA may encode one or more donor polynucleotides.

[0378] A retrotransposon RNA may be capable of binding to a retrotransposon polypeptide. Such retrotransposon RNA may comprise one or more elements for binding to the retrotransposon polypeptide. Examples of binding elements include hairpin structures, pseudoknots (e.g., a nucleic acid secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem), stem loops, and bulges (e.g., unpaired stretches of nucleotides located within one strand of a nucleic acid duplex). In certain examples, the retrotransposon RNA comprises one or more hairpin structures. In some examples, the retrotransposon RNA comprises one or more pseudoknots. In certain examples, a retrotransposon RNA comprises a sequence encoding a donor polynucleotide and one or more binding elements for forming a complex with the retrotransposon polypeptide. The binding elements may be located on the 5' end or the 3' end.

[0379] In one embodiment, a retrotransposon RNA comprises a region capable of hybridizing with an overhang of a target polynucleotide at the target site. The overhang may be a stretch of single-stranded DNA. The overhang may function as a primer for reverse transcription of at least a portion of the retrotransposon RNA to a cDNA. In some cases, a region of the cDNA may be capable of hybridizing a second overhang of the target polynucleotide. The second overhang may function as a primer for the synthesis of a second strand to generate a double-stranded cDNA. The cDNA may comprise a donor polynucleotide sequence. The two overhangs may be from different strands of the target polynucleotide.

[0380] Example CRISPR-Cas, or other programmable nuclease, systems that may be adapted for use with the chimeric IscB systems disclosed herein are disclosed in WO 2021/102042; WO 2022/17380; and Wilkinson *et al.* "Structure of the R2 non-LTR retrotransposon initiating a target-primed reverse transcription" Science. 380(6642), 301-308 (2023), which are incorporated herein by reference in their entirety.

IscB Diversity Generating Retroelement System

[0381] In an embodiment, the chimer IscB systems disclosed herein may further comprise one or more diversity generating retroelement(s) (e.g., DGR described in US20100041033A1). In one embodiment, the DGR may insert a donor polynucleotide with its homing mechanism. For example, the DGR may be associated with a catalytically inactive IscB protein (e.g., a dead IscB), and integrate the single-strand DNA using a homing mechanism. In some examples, the

DGR may be less mutagenic than a counterpart wild type DGR. In some examples, the DGR is not error-prone. In one embodiment, the DGR herein is not mutagenic. The non-mutagenic DGR may be a mutant of a wild type DGR. As used herein, the term "DGR" encompasses both diversity generating retroelement polynucleotides and proteins encoded by diversity generating retroelement polynucleotides. In some examples, DGR may be proteins encoded by diversity generating retroelement polynucleotides having reverse transcriptase activity. In some examples, DGR may be proteins encoded by diversity generating retroelement polynucleotides having reverse transcriptase activity and integrase activity. In some cases, the template or donor polynucleotide may be encoded by a diversity generating retroelement polynucleotide. In certain cases, the template may be a polynucleotide different from the diversity generating retroelement polynucleotide, e.g., provided as a separate construct or molecule.

[0382] In one embodiment, the DGR herein may also include a Group II intron (and any proteins and polynucleotides encoded), which are mobile ribozymes that self-splice from precursor RNAs to yield excised intron lariat RNAs, which then invade new genomic DNA sites by reverse splicing. Examples of Group II intron include those described in Lambowitz AM et al., Group II Introns: Mobile Ribozymes that Invade DNA, Cold Spring Harb Perspect Biol. 2011 Aug; 3(8): a003616.

[0383] In one embodiment, the diversity-generating retroelements (DGRs) are genetic elements that can produce targeted, massive variations in the genomes that carry these elements. In one embodiment, the DGR systems rely on error-prone reverse transcriptases to produce mutagenized cDNA (containing A-to-N mutations) from a template region (TR), to replace a segment called a variable region (VR) that is similar to the TR region—this process is called mutagenic retrohoming (see, e.g., Sharifi and Ye, MyDGR: a server for identification and characterization of diversity-generating retroelements. Nucleic Acids Res. 2019 Jul 2; 47(W1): W289–W294). DGRs may include a unique family of retroelements that generate sequence diversity of DNA. They exist widely in bacteria, archaea, phage and plasmid, and benefit their hosts by introducing variations and accelerating the evolution of target proteins (see, e.g., Yan et al., Discovery and characterization of the evolution, variation and functions of diversity-generating retroelements using thousands of genomes and metagenomes. BMC Genomics. 2019; 20: 595). The first DGR was discovered in a Bordetella phage, BPP-1. *Bordetella* causes the respiratory infection in humans and many other mammals, controlled by the BvgAS signal transduction system. The surface of *Bordetella* is highly variable owing to

the dynamic gene expression in the infectious cycle. The invasion of BPP-1 to *Bordetella* relies on the phage tail fiber protein Mtd. With the process of mutagenic reverse transcription and cDNA integration, DGR may introduce multiple nucleotide substitutions to Mtd gene and generates different receptor-binding molecules, thus making BPP-1 the ability to invade *Bordetellae* with diverse cell surfaces.

[0384] The systems may be used to generate an ssDNA donor using a retron- or DGR RT, which is then integrated by homologous recombination upon target cleavage or nicking using a IscB nuclease. In one embodiment, the systems may comprise DGRs and/or Group-II intron reverse transcriptases. The homing mechanism of DGRs or Group-II introns may be used in modifying a target polynucleotide. The DGRs or Group-II introns reverse transcriptase may be guided to a target polynucleotide by tethering to a nuclease-dead IscB nuclease, TALE, or ZF protein. In another embodiment, a non-retron/DGR reverse transcriptase (e.g. a viral RT) may be used for generating cDNA off of a self-priming RNA. In one embodiment, a ssDNA may be generated by an RT, but integrate it using a dead chimeric IscB system, creating an accessible R-loop instead of nicking/cleaving.

IscB Helitron Systems

[0385] The systems and compositions herein may comprise an chimeric IscB system, one or more ωRNAs or guide RNAs, and one or more components of a helitron. The systems and compositions may be used to insert a donor polynucleotide to a target polynucleotide. The systems and compositions may further comprise a donor polynucleotide.

[0386] The term "helitron", as used herein, refers to a polynucleotide (or nucleic acid segment), recognized as a transposon that captures and mobilizes gene fragments in eukaryotes. The term "helitron" as used herein refers to transposase that comprises an endonuclease domain and a C-terminal helicase domain. Helitrons are rolling-circle RNA transposons. In one embodiment, the helitron encodes a 1400 to about 2000 amino acid, or about 1800 amino acid multidomain transposase. In embodiments, the helitron comprises a hairpin near the 3'end to function as a transposition terminator. In embodiments, the transposon comprises a RepHel motif comprising a replication initiator (Rep) and a DNA helicase (hel) domain. See, Thomas J. & Pritham E. J. Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol. Spectr. 3, 893–926 (2015). In embodiments, the helitron comprises a Rep nuclease domain and C-terminal helicase domain and inserts between an AT dinucleotide in single strand DNA. In an aspect, the C-terminal helicase unwinds the DNA in a 5' to 3' direction. The HUH nuclease

domain may comprise one or two active site tyrosine residues, in embodiments, is a 2 Tyrosine (Y2) HUH endonuclease domain. Helitrons can encompass helentron, proto-helentron and helitron2 type proteins, structures of which can be as described in Thomas et al., 2015 at Figures 1 and 3, incorporated specifically by reference. Particular organisms in which the helitron or helentrons have been found can include those in Table 1 of Thomas J. & Pritham E. J. Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol. Spectr. 3, 893–926 (2015), incorporated herein by reference. Similarly, helitrons can be identified based at least in part on the Rep motif, and conserved residues in the helitrons, and according to the alignment sequence of Figure 2 of Thomas J. & Pritham E. J. Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol. Spectr. 3, 893–926 (2015), specifically incorporated herein by reference.

[0387] The expression "helitron reaction" used herein refers to a reaction wherein a transposase inserts a donor polynucleotide sequence in or adjacent to an insertion site on a target polynucleotide. The insertion site may contain a sequence or secondary structure recognized by the helitron and/or an insertion motif sequence in the target polynucleotide into which the donor polynucleotide sequence may be inserted.

[0388] As described in Grabundzija 2018, the helitron terminal sequences contains a distinct ~150 base pairs (bp) long sequence with an absolutely conserved dinucleotide at the end of left terminal sequence (LTS), and a tetranucleotide at the end of right terminal sequence (RTS) which is preceded by a palindromic sequence that can form a hairpin structure. Grabundzija et al., Nat. Commun. 2018; 9: 1278; doi:10.1035/s41467-018-03688-w.

[0389] The helitron end sequences may be responsible for identifying the donor polynucleotide for transposition. The helitron end sequences may be the DNA sequences used to perform a transposition reaction, the end sequences may be referred to herein as right terminal sequences and left terminal sequence. The donor polynucleotide can be configured to comprise a first and second helitron recognition sequence that are at least 80%, 85%, 90%, 95% 96%, 97%, 98%, 99% or 100% complementary to a left terminal sequence and/or a right terminal sequence of a polynucleotide encoding the helitron polypeptide.

[0390] In an aspect, the palindromic sequence may be located upstream of the right terminal sequence, for example, about 5, 10, 15, 20, 25, 30, 35 nucleotides upstream of the right terminal sequence end, or about 10 to 15 nucleotides upstream of the right terminal sequence end, about 10 to 12 nucleotides or about 11 nucleotides upstream of the right terminal

sequence end. Ivana Grabundzija, Nat Commun. 2016; 7:10716, doi:10.1038/ncomms10716, incorporated herein by reference.

[0391] Exemplary helitrons can be identified using software, for example (EAHelitron) that has been used to identify Helitrons in a wide range of plant genomes. See, Hu, K., Xu, K., Wen, J. *et al. Helitron* distribution in Brassicaceae and whole Genome Helitron density as a character for distinguishing plant species. *BMC Bioinformatics* 20, 354 (2019). doi: 10.1186/s12859-019-2945-8, incorporated herein by reference.

[0392] The helitron may be derived from a eukaryote. In an aspect, the helitron is derived from a mammalian genome, in an aspect, vespertilionid bats, e.g., Helibat. In embodiments, the helitron is derived from derived from a Helibat1 transposon. In embodiments, the helitron is Helraiser, the full DNA sequence of the consensus transposon, including left terminal and right terminal sequences as well as hairpin identified is provided in Grabundzija, 2016 at Supplementary Figure 1, specifically incorporated herein by reference. In an aspect, the helitron is flanked by left and right terminal sequences of the transposon. In an aspect, the left terminal sequence and right terminal sequence terminates with the conserved 5′-TC/CTAG-3′ motif. In an embodiment, the helitron may comprise a palindromic sequence that is about 10 to about 35, or about 5-25 bp or about 19-bp-long palindromic sequence with the potential to form a hairpin structure.

[0393] Elements of these systems may be engineered to work within the context of the invention. For example, a helitron polypeptide may be fused to a polypeptide capable of generating an R-loop. Fusion may be by any appropriate linker, in an exemplary embodiment, XTEN16. The binding elements that allow a helitron polypeptide to bind, for example, the use of sequences complementary to the right terminal sequence and the left terminal sequence of the helitron may be engineered into a donor construct to facilitate entry of a donor polynucleotide sequence into a target polynucleotide.

[0394] In certain example embodiments, the Isc polypeptide, via formation of complex with a ωRNA sequence, directs the helitron polypeptide to a target sequence in a target polynucleotide, where the helitron facilitates integration of a donor polynucleotide sequence into the target polynucleotide.

[0395] The helitron polypeptides may also comprise one or more truncations or excisions to remove domains or regions of wild-type protein to arrive at a minimal polypeptide, alter functionality according to the system in which the helitron is used, or mutated to enhance or

diminish particular activities associated with the helitron, i.e., nuclease activity or helicase activity.

IscB Recombinase/Integrase Systems

[0396] The systems and compositions herein may comprise a chimeric IscB system, and one or more components of a recombinase or integrase. In an aspect, the chimeric IscB system is naturally catalytically inactive and utilized with one or more nucleic acid components to provide site-specific targeting, and the one or more components of the recombinase to introduce a modification. In an aspect, the chimeric IscB system may be catalytically inactivated via mutation of one or more residues of a catalytic domain or via truncation and utilized with one or more RNA components to provide site-specific targeting, and the one or more components of the recombinase introduce a modification. In one embodiment, a naturally inactive IscB is provided with a recombinase, e.g., an integrase, and optionally a reverse transcriptase.

[0397] A recombinase generally is an enzyme that mediates recombination, e.g., breaking and rejoining, of nucleic acids at specific points. DNA site-specific recombinases include serine integrases, which are phage-encoded site-specific recombinases that promote conservative recombination reactions between DNA substrates located on the phage (phage attachment site, *attP*) and bacterial attachment site, *attB*. In one embodiment, the recombinase is a serine integrase that drives a highly directions site-specific recombination.

[0398] In preferred embodiments, the recombinase mediates unidirectional site-specific recombination. In one embodiment, the recombinase is a serine recombinase (SR) also referred to as a serine integrase, encoded, for example, by IS607 family, Tn4451, and bacteriophage phiC31. See, generally, Smith MC, Thorpe HM: Diversity in the serine recombinases. Mol Microbiol. 2002, 44: 299-307. 10.1046/j.1365-2958.2002.02891.x; Li et al.,(2018) J. Mol. Biol. 430:21, 4401-4418.

[0399] In an embodiment, the recombinase is a tyrosine recombinase (YR) encoded by IS91, Helitron, IS200/IS605, Crypton or DIRS-retrotransposon families. See, generally, Goodwin TJ, Butler MI, Poulter T: Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. Microbiology. 2003, 149: 3099-3109. Doi:10.1099/mic.0.26529-0; Cappello J, Handelsman K, Lodish HF: Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell. 1985, 43: 105-115. 10.1016/0092-8674(85)90016-9.

[0400] In an aspect, the recombinase provides site-specific integration of a template that can be provided with the composition, e.g., a donor oligonucleotide. Without being bound by theory, the recombinase allows for integration independent of payload size and can coordinate strand exchange and re-ligation across multiple cell types, allowing integration of long stretches of polynucleotides. In an exemplary embodiment, the serine recombinase is PhiC31 and the target is DNA. In an aspect, the phiC31 allows for integration of a target site comprising an attP or pseudoattP recognition site. See, e.g., systembio.com/wp-content/uploads/phiC31_productsheet-1.pdf. In an embodiment utilizing phiC231, a donor oligonucleotide would be provided with an attB at sequence that facilitates attachment at the attP site of the target genome. Similar approaches of designing donor oligonucleotides with sequences complementary to attachment sites for a recombinase can be designed for use with the present invention. See, e.g. Li et al.,(2018) J. Mol. Biol. 430:21, 4401-4418.

[0401] In preferred embodiments, the integrase mediates gene integration at diverse loci by directing insertion with an IscB nickase fused to both a reverse transcriptase and an integrase. In one embodiment, the integrase is a serine integrase, encoded, for example, BxbINT. See, generally, Ioannidi et al., "Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed integrases"; doi:10.1101/2021.11.01.466786m incorporated herein by reference in its entirety. In Ioannidi, Gootenberg, Abudayyeh, and colleagues show integration using a CRISPR-Cas9 nickase fused to a reverse transcriptase and serine integrase termed Programmable Addition via Site-specific Targeting Elements (PASTE) with delivery via a single dose of plasmids with functionality in non-dividing and primary cells, utilizing a guide RNA comprising an AttB landing site, termed attachment site-containing guide RNA were used to insert sequences, including diverse cargo sequences that can be inserted across different loci, varying in size up to about 36 kb. Additional uses of the PASTE system included gene tagging, gene replacement, gene delivery, and protein production and secretion, approaches that are contemplated for use with the IscB nickase and integrase approach. In an aspect, the ωRNA may comprise an AttB landing site. In an aspect, the recombinase provides site-specific integration of a template that can be provided with the composition, e.g., a donor oligonucleotide.

[0402] Additional large serine integrases can be used with the IscB nickase, for example as identified and described in Durrant et al., Large-scale discovery of recombinases for integrating DNA into the human genome, doi:10.1101/2021.11.05.467528, incorporated herein

by reference. Other integrases include BceINT, SscINT, SacINT. See, Ioannidi, 2021 at and Fig. 6d, and Fig. 10a.

[0403] Without being bound by theory, the recombinase allows for integration independent of payload size and can coordinate strand exchange and re-ligation across multiple cell types, allowing integration of long stretches of polynucleotides. In an exemplary embodiment, the integrase is BxbINT and the target is DNA. In an aspect, the BxbINT allows for integration of a target site comprising an attP or pseudoattP recognition site. In an embodiment utilizing BxbINT, a donor oligonucleotide would be provided with an attB at sequence that facilitates attachment at the attP site of the target genome. Similar approaches of designing donor oligonucleotides with sequences complementary to attachment sites for an integrase can be designed for use with the present invention, for example a circular double-strand DNA template containing the AttP attachment site, or delivery of large cargo via an adenovirus or other viral vector, as described elsewhere herein. See, e.g., Ioannidi et al., 2021 at Fig 1a, 1b and 5b.

IscB Topoisomerase Systems

[0404] The one or more functional domains may be one or more topoisomerase domains. In one embodiment, an engineered system for modifying a target polynucleotide comprising: an chimeric IscB system; a topoisomerase domain; and a nucleic acid template comprising or encoding a donor polynucleotide to be inserted to a target sequence of the target polynucleotide. In some examples, two or more of: the chimeric IscB system; topoisomerase domain; and nucleic acid template may form a complex. In some examples, two or more of: the chimeric IscB system; topoisomerase domain, may be comprised in a fusion protein.

[0405] Topoisomerases are a class of enzymes that modify the topological state of DNA via the breakage and rejoining of nucleic acid strands. In some cases, a topoisomerase may be a DNA topoisomerase, which is an enzyme that controls and alters the topologic states of DNA during transcription, and catalyzes the transient breaking and rejoining of a single strand of DNA which allows the strands to pass through one another, thus altering the topology of DNA.

[0406] In one embodiment, the topoisomerase domain is capable of ligating the donor polynucleotide with the target polynucleotide. The ligation may be achieved by sticky end or blunt end ligation. In an example, the donor polynucleotide may comprise an overhang comprising a sequence complementary to a region of the target polynucleotide. Examples of ligating the donor polynucleotide with the target polynucleotide include those of TOPO cloning, e.g., those described in "The Technology Behind TOPO Cloning," at

www.thermofisher.com/us/en/home/life-science/cloning/topo/topo-resources/the-technology-behind-topo-cloning.html.

[0407] In one embodiment, the topoisomerase domain may be associated with the donor polynucleotide. For example, the topoisomerase domain is covalently linked to the donor polynucleotide.

[0408] In one embodiment, a topoisomerase domain may be provided together with, e.g., associated (e.g., fused) with a chimeric IscB system (e.g., a chimeric IscB system or a variant thereof such as a dead IscB or a IscB nickase). Alternatively or additionally, the topoisomerase domain may be on a molecule different from the chimeric IscB system. In some cases, the topoisomerase domain may be associated with a donor polynucleotide. For example, the topoisomerase domain may be pre-loaded covalently with a donor DNA molecule. Such design may allow for efficient ligation of only a specific cargo. The topoisomerase domain may ligate the donor polynucleotide (e.g., a DNA molecule) to a target site on a target polynucleotide (e.g., a free double-stranded DNA end). In one embodiment, the donor polynucleotide may have an overhang that comprises a sequence complementary to a region of the target polynucleotide. For example, the overhang may invade into the target polynucleotide at a cut site generated by the chimeric IscB system.

[0409] Examples of topoisomerases include type I, including type IA and type IB topoisomerases, which cleave a single strand of a double-stranded nucleic acid molecule, and type II topoisomerases (e.g., gyrases), which cleave both strands of a double-stranded nucleic acid molecule.

[0410] Type IA and IB topoisomerases cleave one strand of a double-stranded nucleic acid molecule. In some examples, the cleavage of a double-stranded nucleic acid molecule by type IA topoisomerases generates a 5′ phosphate and a 3′ hydroxyl at the cleavage site, with the type IA topoisomerase covalently binding to the 5′ terminus of a cleaved strand. Cleavage of a double-stranded nucleic acid molecule by type IB topoisomerases may generate a 3′ phosphate and a 5′ hydroxyl at the cleavage site, with the type IB topoisomerase covalently binding to the 3′ terminus of a cleaved strand.

[0411] Examples of Type IA topoisomerases include *E. coli* topoisomerase I, *E. coli* topoisomerase III, eukaryotic topoisomerase II, archeal reverse gyrase, yeast topoisomerase III, Drosophila topoisomerase III, human topoisomerase III, Streptococcus pneumoniae topoisomerase III, and the like, including other type IA topoisomerases. A DNA-protein adduct

is formed with the enzyme covalently binding to the 5′-thymidine residue, with cleavage occurring between the two thymidine residues.

[0412] Examples of Type IB topoisomerases include the nuclear type I topoisomerases present in all eukaryotic cells and those encoded by Vaccinia and other cellular poxviruses. The eukaryotic type IB topoisomerases are exemplified by those expressed in yeast, *Drosophila* and mammalian cells, including human cells. Viral type IB topoisomerases are exemplified by those produced by the vertebrate poxviruses (Vaccinia, Shope fibroma virus, ORF virus, fowlpox virus, and molluscum contagiosum virus), and the insect poxvirus (*Amsacta moorei entomopoxvirus*).

[0413] Examples of Type II topoisomerases include, bacterial gyrase, bacterial DNA topoisomerase IV, eukaryotic DNA topoisomerase II, and T-even phage encoded DNA topoisomerases. Type II topoisomerases may have both cleaving and ligating activities. Substrate double-stranded nucleic acid molecules of type II topoisomerase can be prepared such that the type II topoisomerase can form a covalent linkage to one strand at a cleavage site. For example, calf thymus type II topoisomerase can cleave a substrate ds nucleic acid molecule containing a 5′ recessed topoisomerase recognition site positioned three nucleotides from the 5′ end, resulting in dissociation of the three nucleic acid molecule 5′ to the cleavage site and covalent binding of the topoisomerase to the 5′ terminus of the ds nucleic acid molecule. Furthermore, upon contacting such a type II topoisomerase-charged ds nucleic acid molecule with a second nucleic acid molecule containing a 3′ hydroxyl group, the type II topoisomerase can ligate the sequences together, and then is released from the recombinant nucleic acid molecule.

[0414] In some examples, the topoisomerase is DNA topoisomerase I, e.g., a Vaccinia virus topoisomerase I. The topoisomerase may be pre-loaded with a donor polynucleotide. The Vaccinia virus topoisomerase may need a target comprising a 5’ -OH group.

IscB Phosphatase System

[0415] The systems herein may further comprise a phosphatase domain. A phosphatase is an enzyme capable of removing a phosphate group from a molecule e.g., a nucleic acid such as DNA. Examples of phosphatases include calf intestinal phosphatase, shrimp alkaline phosphatase, Antarctic phosphatase, and APEX alkaline phosphatase.

[0416] In some examples, the 5’ -OH group of in the target polynucleotide may be generated by a phosphatase. A topoisomerase compatible with a 5' phosphate target may be

used to generate stable loaded intermediates. In some cases, a chimeric IscB system that leaves a 5' OH after cleaving the target polynucleotide may be used. In some cases, the phosphatase domain may be associated with (e.g., fused to) the IscB protein. The phosphatase domain may be capable of generating a -OH group at a 5' end of the target polynucleotide. The phosphatase may be delivered separated from other components in the system, e.g., as a separate protein, on a separate vector from other components.

IscB Polymerase System

[0417] The systems herein may further comprise a polymerase domain. A polymerase refers to an enzyme that synthesizes chains of nucleic acids. The polymerase may be a DNA polymerase or an RNA polymerase.

[0418] In one embodiment, the systems comprise an engineered system for modifying a target polynucleotide comprising: a chimeric IscB system; a DNA polymerase domain; and a DNA template comprising a donor polynucleotide to be inserted to a target sequence of the target polynucleotide. In some examples, two or more of: the IscB protein; DNA polymerase domain; and DNA template may form a complex. In some examples, two or more of: the IscB protein; DNA polymerase domain; are comprised in a fusion protein. For example, the chimeric IscB system and DNA polymerase domain may be comprised in a fusion protein.

[0419] In one embodiment, the systems may comprise a chimeric IscB system (or variant thereof such as a dIscB polypeptide or chimeric IscB system) and a DNA polymerase (e.g., phi29, T4, T7 DNA polymerase). The systems may further comprise a single-stranded DNA or double-stranded DNA template. The DNA template may comprise i) a first sequence homologous to a target site of the IscB protein on the target polynucleotide, and/or ii) a second sequence homologous to another region of the target polynucleotide. In one embodiment, the template may be a synthetic single-stranded or PCR-generated DNA molecule, (optionally end-protected by modified nucleotides), or a viral genome (e.g., AAV). In another embodiment, the template is generated using a reverse transcriptase. When the system is delivered into a cell, an endogenous DNA polymerase in the cell may be used. Alternatively or additionally, an exogenous DNA polymerase may be expressed in the cell.

[0420] The DNA template may be end-protected by one or more modified nucleotides, or comprises a portion of a viral genome. In some embodiment, the DNA template comprises LNA or other modifications (e.g., at the 3' end). The presence of LNA and/or the modifications

may lead to more efficient annealing with the 3' flap generated by chimeric IscB system cleavage.

[0421] Examples of DNA polymerase include Taq, Tne (exo −), Tma (exo −), Pfu (exo −), Pwo (exo −), Thermoanaerobacter thermohydrosulfuricus DNA polymerase, Thermococcus litoralis DNA polymerase I, E. coli DNA polymerase I, Taq DNA polymerase I, Tth DNA polymerase I, Bacillus stearothermophilus (Bst) DNA polymerase I, E. coli DNA polymerase III, bacteriophage T5 DNA polymerase, bacteriophage M2 DNA polymerase, bacteriophage T4 DNA polymerase, bacteriophage T7 DNA polymerase, bacteriophage phi29 DNA polymerase, bacteriophage PRD1 DNA polymerase, bacteriophage phi15 DNA polymerase, bacteriophage phi21DNA polymerase, bacteriophage PZE DNA polymerase, bacteriophage PZA DNA polymerase, bacteriophage Nf DNA polymerase, bacteriophage M2Y DNA polymerase, bacteriophage B103 DNA polymerase, bacteriophage SF5 DNA polymerase, bacteriophage GA-1 DNA polymerase, bacteriophage Cp-5 DNA polymerase, bacteriophage Cp-7 DNA polymerase, bacteriophage PR4 DNA polymerase, bacteriophage PR5 DNA polymerase, bacteriophage PR722 DNA polymerase and bacteriophage L17 DNA polymerase.

IscB Reverse Transcriptase Systems

[0422] The one or more functional domains may comprise alone, or in additional to additional functional domains, one or more reverse transcriptase domains. In one embodiment, the systems comprise an engineered system for modifying a target polynucleotide comprising: a chimeric IscB system or a variant thereof (e.g., dIscB); a reverse transcriptase (RT) domain; a RNA template comprising or encoding a donor polynucleotide to be inserted to a target sequence of the target polynucleotide; and an ωRNA or guide RNA molecule (i.e., a naturally single guide RNA molecule comprising a scaffold for reprogramming).

[0423] The reverse transcriptase may generate single-strand DNA based on the RNA template. The single-strand DNA may be generated by a non-retron, retron, or diversity generating retroelement (DGR). In some examples, the single-strand DNA may be generated from a self-priming RNA template. A self-priming RNA template may be used to generate a DNA without the need of a separate primer.

[0424] A reverse transcriptase domain may be a reverse transcriptase or a fragment thereof. A wide variety of reverse transcriptases (RT) may be used in alternative embodiments of the present invention, including prokaryotic and eukaryotic RT, provided that the RT functions within the host to generate a donor polynucleotide sequence from the RNA template. If desired,

the nucleotide sequence of a native RT may be modified, for example using known codon optimization techniques, so that expression within the desired host is optimized. A reverse transcriptase (RT) is an enzyme used to generate complementary DNA (cDNA) from an RNA template, a process termed reverse transcription. Reverse transcriptases are used by retroviruses to replicate their genomes, by retrotransposon mobile genetic elements to proliferate within the host genome, by eukaryotic cells to extend the telomeres at the ends of their linear chromosomes, and by some non-retroviruses such as the hepatitis B virus, a member of the Hepadnaviridae, which are dsDNA-RT viruses. Retroviral RT has three sequential biochemical activities: RNA-dependent DNA polymerase activity, ribonuclease H, and DNA-dependent DNA polymerase activity. Collectively, these activities enable the enzyme to convert single-stranded RNA into double-stranded cDNA. In an embodiment, the RT domain of a reverse transcriptase is used in the present invention. The domain may include only the RNA-dependent DNA polymerase activity. In some examples, the RT domain is non-mutagenic, i.e., does not cause mutation in the donor polynucleotide (e.g., during the reverse transcriptase process). In some examples, the RT domain may be non-retron RT, e.g., a viral RT or a human endogenous RT. In some examples, the RT domain may be retron RT or DGRs RT. In some examples, the RT may be less mutagenic than a counterpart wildtype RT. In one embodiment, the RT herein is not mutagenic.

Retrons

[0425] In an embodiment, a donor template for homologous recombination is generated by use of a self-priming RNA template for reverse transcription. A non-limiting example of a self-priming reverse transcription system is the retron system. By the term “retron” it is meant a genetic element which encodes components enabling the synthesis of branched RNA-linked single stranded DNA (msDNA) and a reverse transcriptase. Retrons which encode msDNA are known in the art, for example, but not limited to U.S. Pat. No. 6,017,737; U.S. Pat. No. 5,849,563; U.S. Pat. No. 5,780,269; U.S. Pat. No. 5,436,141; U.S. Pat. No. 5,405,775; U.S. Pat. No. 5,320,958; CA 2,075,515; all of which are herein incorporated by reference.

[0426] In an embodiment, the reverse transcriptase domain is a retron RT domain. In an embodiment, the RNA template encodes a retron RNA template that is recognized and reverse transcribed by the retron reverse transcriptase domain. Conserved across many bacterial species, retrons are highly efficient reverse transcription systems of relatively unknown function. The retron system consists of the retron RT protein, as well as the msr and msd

transcripts, which function as the primer and template sequences, respectively. All components of the retron system are expressed from a single open reading frame as a single transcript including the msr-msd and encoding the retron RT protein (Lampson, et al., 2005, Retrons, msDNA, and the bacterial genome. Cytogenet Genome Res 110:491–499). The msr element ORF of a retron provides for the RNA portion of the msDNA molecule, while the msd element ORF provides for the DNA portion of the msDNA molecule. The primary transcript from the msr-msd region is thought to serve as both a template and a primer to produce the msDNA. Synthesis of msDNA is primed from an internal rG residue of the RNA transcript using its 2′-OH group. Modification of msd, or msr may also be made to permit insertion of a RNA template encoding a donor polynucleotide within the msd without altering the functioning of or the production of msDNA. The RNA template encoding a donor polynucleotide sequence may be any length but is preferably less than about 5 kb nucleotides, or also less than about 2 kb, or also less than 500 bases, provided that an msDNA product is produced.

IscB Ligase Systems

[0427] In general, the systems comprise a chimeric IscB system and a ligase associated with the IscB protein. The chimeric IscB system, may be recruited to the target sequence by an ωRNA, or guide RNA, and generate a break on the target sequence. The ωRNA or guide RNA may further comprise a template sequence with desired mutations or other sequence elements. The template sequence may be ligated to the target sequence to introduce the mutations or other sequence elements to the nucleic acid molecule. The chimeric IscB system, may be a nickase that generates a single-strand break on nucleic acid molecule, and the ligase may be a single-strand DNA ligase. In one embodiment, the systems comprise a pair of chimeric IscB system complexes, with two distinct ωRNA (or guide) sequences. Each chimeric IscB system complex can target one strand of a double-stranded polynucleotide, and work together to effectively modify the sequence of the double-stranded polynucleotides.

[0428] In some examples, the chimeric IscB system is associated with a ligase or functional fragment thereof. The ligase may ligate a single-strand break (a nick) generated by the chimeric IscB system. In certain cases, the ligase may ligate a double-strand break generated by the chimeric IscB system. In certain examples, the chimeric IscB system is associated with a reverse transcriptase or functional fragment thereof.

[0429] The present invention further provides systems and methods of modifying a nucleic acid sequence using a pair of distinct chimeric IscB system-ligase-ωRNA or guide RNA

complexes, said systems and methods comprising: (a) an engineered chimeric IscB system connected to or complexed with a ligase; (b) two distinct ωRNA or guide RNA sequences complexed with such chimeric IscB system-ligase protein complex to form a first and a second distinct IscB-ligase ωRNA complexes; (c) the first IscB-ligase–ωRNA or guide RNA complex binding to one strand of a target double-stranded polynucleotide sequence, and the second chimeric IscB system-ligase- ωRNA or guide RNA complex binding to another strand of the target double-stranded polynucleotide sequence; (d) upon binding of the said complexes to the locus of interest the effector protein induces the modification of the sequences associated with or at the target locus of interest, whereby the two chimeric IscB system-ligase- ωRNA or guide RNA complexes work together on different strands of the double-stranded target sequence and modify the sequence.

[0430] One of the advantages of using such a "pair" of chimeric IscB system-ligase-ωRNA or guide RNA complexes includes high efficiency in modifying the sequence associated with or at the locus of interest of target double-stranded polynucleotides.

[0431] In one embodiment, the chimeric IscB system can be a nickase. In a preferred embodiment, a ligase is linked to the chimeric IscB system. The ligase can ligate the donor sequence to the target sequence. The ligase can be a single-strand DNA ligase or a double-strand DNA ligase. The ligase can be fused to the carboxyl-terminus of a chimeric IscB system, or to the amino-terminus of a chimeric IscB system.

[0432] As used herein the term "ligase" refers to an enzyme, which catalyzes the joining of breaks (e.g., double-stranded breaks or single-stranded breaks ("nicks") between adjacent bases of nucleic acids. For example, a ligase may be an enzyme capable of forming intra- or inter-molecular covalent bonds between a 5′ phosphate group and a 3′ hydroxyl group. The term "ligate" refers to the reaction of covalently joining adjacent oligonucleotides through formation of an internucleotide linkage.

[0433] DNA ligases fall into two general categories: ATP-dependent DNA ligases (EC 6.5.1.1), and NAD (+) dependent DNA ligases (EC 6.5.1.2). NAD (+) dependent DNA ligases are found only in bacteria (and some viruses) while ATP-dependent DNA ligases are ubiquitous. The ATP-dependent DNA ligases can be divided into four classes: DNA ligase I, II, III, and IV. DNA ligase I links Okazaki fragments to form a continuous strand of DNA; DNA ligase II is an alternatively spliced form of DNA ligase III, found only in non-dividing cells; DNA ligase III is involved in base excision repair; and DNA ligase IV is involved in the

repair of DNA double-strand breaks by non-homologous end joining (NHEJ). Amongst all ligases, there are two types of prokaryotic and one type of eukaryotic ligases that are particularly well suited for facilitating the blunt-ended, double-stranded DNA ligation: Prokaryotic DNA ligases (T3 and T4) and Eukaryotic DNA ligase (Ligase 1).

[0434] In some cases, the ligase is specific for double-stranded nucleic acids (e.g., dsDNA, dsRNA, RNA/DNA duplex). An example of a ligase specific for double-stranded DNA and DNA/RNA hybrids is T4 DNA ligase. In some cases, the ligase is specific for single-stranded nucleic acids (e.g., ssDNA, ssRNA). An example of such ligase is CircLigase II. In some cases, the ligase is specific for RNA/DNA duplexes. In some cases, the ligase is able to work on single-stranded, double-stranded, and/or RNA/DNA nucleic acids in any combination.

[0435] In some cases, the ligase may be a pan-ligase, which is a single ligase with the ability to ligate both DNA and RNA targets. The ligase may be specific for a target (e.g., DNA-specific or RNA-specific). In some cases, the ligase may be a dual ligase system that include DNA-specific, RNA-specific, and/or pan-ligases, in any combination.

[0436] Examples of ligases that can be used with the disclosure include T4 DNA Ligase, T3 DNA Ligase, T7 DNA Ligase, E. coli DNA Ligase, HiFi Taq DNA Ligase, 9° N™ DNA Ligase, Taq DNA Ligase, SplintR® Ligase (also known as. PBCV-1 DNA Ligase or Chlorella virus DNA Ligase), Thermostable 5′ AppDNA/RNA Ligase, T4 RNA Ligase, T4 RNA Ligase 2, T4 RNA Ligase 2 Truncated, T4 RNA Ligase 2 Truncated K227Q, T4 RNA Ligase 2, Truncated KQ, RtcB Ligase (joins single stranded RNA with a 3″-phosphate or 2′,3′-cyclic phosphate to another RNA), CircLigase II, CircLigase ssDNA Ligase, CircLigase RNA Ligase, or Ampligase® Thermostable DNA Ligas, NAD-dependent ligases including Taq DNA ligase, Thermus filiformis DNA ligase, Escherichia coliDNA ligase, Tth DNA ligase, Thermus scotoductus DNA ligase (I and II), thermostable ligase, Ampligase thermostable DNA ligase, VanC-type ligase, 9° N DNA Ligase, Tsp DNA ligase, and novel ligases discovered by bioprospecting; ATP-dependent ligases including T4 RNA ligase, T4 DNA ligase, T3 DNA ligase, T7 DNA ligase, Pfu DNA ligase, DNA ligase I, DNA ligase III, DNA ligase IV, and novel ligases discovered by bioprospecting, and wild-type, mutant isoforms, and genetically engineered variants thereof.

[0437] In one embodiment, the examples of the ligases include those used in sequencing by synthesis or sequencing by ligation reactions.

Iscb Epigenetic Editors

[0438] In one embodiments, the chimeric IscB systems described herein may further comprise an epigenetic modification domain such that binding of the chimeric IscB at target sequence on genomic DNA (e.g., chromatin) results in one or more epigenetic modifications by the epigenetic modification domain that increases or decreases expression of the one or more polypeptides. As used herein, "linked to or otherwise capable of associating with" refers to a fusion protein or a recruitment domain or an adaptor protein, such as an aptamer (e.g., MS2) or an epitope tag. The recruitment domain or an adaptor protein can be linked to an epigenetic modification domain or the DNA binding domain (e.g., an adaptor for an aptamer). The epigenetic modification domain can be linked to an antibody specific for an epitope tag fused to the chimeric IscB. An aptamer can be linked to a guide sequence.

[0439] In example embodiments, the DNA binding domain is a programmable DNA binding protein linked to or otherwise capable of associating with an epigenetic modification domain. In example embodiments, the DNA binding domain is a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient endonuclease enzyme. In example embodiments, a CRISPR system having an inactivated nuclease activity (e.g., dCas) is used as the DNA binding domain.

[0440] In example embodiments, the epigenetic modification domain is a functional domain and includes, but is not limited to a histone methyltransferase (HMT) domain, histone demethylase domain, histone acetyltransferase (HAT) domain, histone deacetylation (HDAC) domain, DNA methyltransferase domain, DNA demethylation domain, histone phosphorylation domain (e.g., serine and threonine, or tyrosine), histone ubiquitylation domain, histone sumoylation domain, histone ADP ribosylation domain, histone proline isomerization domain, histone biotinylation domain, histone citrullination domain (see, e.g., Epigenetics, Second Edition, 2015, Edited by C. David Allis; Marie-Laure Caparros; Thomas Jenuwein; Danny Reinberg; Associate Editor Monika Lachlan; Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12-27; Syding LA, Nickl P, Kasparek P, Sedlacek R. CRISPR/Cas9 Epigenome Editing Potential for Rare Imprinting Diseases: A Review. Cells. 2020;9(4):993; and Zhang Y. Transcriptional regulation by histone ubiquitination and deubiquitination. Genes Dev. 2003;17(22):2733-2740). Example epigenetic modification domains can be obtained from, but are not limited to chromatin modifying enzymes, such as, DNA methyltransferases (e.g., DNMT1, DNMT3a and DNMT3b), TET1,

TET2, thymine-DNA glycosylase (TDG), GCN5-related N-acetyltransferases family (GNAT), MYST family proteins (e.g., MOZ and MORF), and CBP/p300 family proteins (e.g., CBP, p300), Class I HDACs (e.g., HDAC 1-3 and HDAC8), Class II HDACs (e.g., HDAC 4-7 and HDAC 9-10), Class III HDACs (e.g., sirtuins), HDAC11, SET domain containing methyltransferases (e.g., SET7/9 (KMT7, NCBI Entrez Gene: 80854), KMT5A (SET8), MMSET, EZH2, and MLL family members), DOT1L, LSD1, Jumonji demethylases (e.g., KDM5A (JARID1A), KDM5C (JARID1C), and KDM6A (UTX)), kinases (e.g., Haspin, VRK1, PKCα, PKCβ, PIM1, IKKα, Rsk2, PKB/Akt, Aurora B, MSK1/2, JNK1, MLTKα, PRK1, Chk1, Dlk/ZIP, PKCδ, MST1, AMPK, JAK2, Abl, BMK1, CaMK, S6K1, SIK1), Ubp8, ubiquitin C-terminal hydrolases (UCH), the ubiquitin-specific processing proteases (UBP), and poly(ADP-ribose) polymerase 1 (PARP-1). See, also, US Patent US11001829B2 for additional domains. In example embodiments, the epigenetic modification domain is a catalytically active IscB polypeptide described herein.

[0441] In example embodiments, histone acetylation is targeted to a target sequence using a chimeric IscB polypeptide (see, e.g., Hilton IB, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol. 2015). In example embodiments, histone deacetylation is targeted to a target sequence (see, e.g., Cong et al., 2012; and Konermann S, et al. Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013;500:472–476). In example embodiments, histone methylation is targeted to a target sequence (see, e.g., Snowden AW, Gregory PD, Case CC, Pabo CO. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Curr Biol. 2002;12:2159–2166; and Cano-Rodriguez D, Gjaltema RA, Jilderda LJ, et al. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat Commun. 2016;7:12284). In example embodiments, histone demethylation is targeted to a target sequence (see, e.g., Kearns NA, Pham H, Tabak B, et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods. 2015;12(5):401-403). In example embodiments, histone phosphorylation is targeted to a target sequence (see, e.g., Li J, Mahata B, Escobar M, et al. Programmable human histone phosphorylation and gene activation using a CRISPR/Cas9-based chromatin kinase. Nat Commun. 2021;12(1):896). In example embodiments, DNA methylation is targeted to a target sequence (see, e.g., Rivenbark AG, et al. Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics. 2012;7:350–360; Siddique AN, et al. Targeted methylation and

gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. J Mol Biol. 2013;425:479–491; Bernstein DL, Le Lay JE, Ruano EG, Kaestner KH. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J Clin Invest. 2015;125:1998–2006; Liu XS, Wu H, Ji X, et al. Editing DNA Methylation in the Mammalian Genome. Cell. 2016;167(1):233-247.e17; Stepper P, Kungulovski G, Jurkowska RZ, et al. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic Acids Res. 2017;45(4):1703-1713; and Pflueger C., Tan D., Swain T., Nguyen T., Pflueger J., Nefzger C., Polo J.M., Ford E., Lister R. A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. Genome Res. 2018;28:1193–1206). In example embodiments, DNA demethylation is targeted to a target sequence using a CRISPR system (see, e.g., TET1, see Xu et al, Cell Discov. 2016 May 3;2: 16009; Choudhury et al, Oncotarget. 2016 Jul 19;7(29):46545-46556; and Kang JG, Park JS, Ko JH, Kim YS. Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. Sci Rep. 2019;9(1):11960). In example embodiments, DNA demethylation is targeted to a target sequence (see, e.g., TDG, see, Gregory DJ, Zhang Y, Kobzik L, Fedulov AV. Specific transcriptional enhancement of inducible nitric oxide synthase by targeted promoter demethylation. Epigenetics. 2013;8:1205–1212).

[0442] Example epigenetic modification domains can be obtained from, but are not limited to transcription activators, such as, VP64 (see, e.g., Ji Q, et al. Engineered zinc-finger transcription factors activate OCT4 (POU5F1), SOX2, KLF4, c-MYC (MYC) and miR302/367. Nucleic Acids Res. 2014;42:6158–6167; Perez-Pinera P, et al. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat Methods. 2013;10:239–242; Farzadfard F, Perli SD, Lu TK. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. ACS Synth Biol. 2013;2:604–613; Black JB, Adler AF, Wang HG, et al. Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell. 2016;19(3):406-414; and Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977-979), p65 (see, e.g., Liu PQ, et al. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of

vascular endothelial growth factor A. J Biol Chem. 2001;276:11323–11334; and Konermann S, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015;517:583–588), HSF1, and RTA (see, e.g., Chavez A, et al. Highly efficient Cas9-mediated transcriptional programming. Nat Methods. 2015;12:326–328). Example epigenetic modification domains can be obtained from, but are not limited to transcription repressors, such as, KRAB (see, e.g., Beerli RR, Segal DJ, Dreier B, Barbas CF., 3rd Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A. 1998;95:14628–14633; Cong L, Zhou R, Kuo YC, Cunniff M, Zhang F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun. 2012;3:968; Gilbert LA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154:442–451; and Yeo NC, Chavez A, Lance-Byrne A, et al. An enhanced CRISPR repressor for targeted mammalian gene regulation. Nat Methods. 2018;15(8):611-616).

[0443] In example embodiments, the epigenetic modification domain linked to a DNA binding domain recruits an epigenetic modification protein to a target sequence. In example embodiments, a transcriptional activator recruits an epigenetic modification protein to a target sequence. For example, VP64 can recruit DNA demethylation, increased H3K27ac and H3K4me. In example embodiments, a transcriptional repressor protein recruits an epigenetic modification protein to a target sequence. For example, KRAB can recruit increased H3K9me3 (see, e.g., Thakore PI, D'Ippolito AM, Song L, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat Methods. 2015;12(12):1143-1149). In an example embodiment, methyl-binding proteins linked to a DNA binding domain, such as MBD1, MBD2, MBD3, and MeCP2 recruits an epigenetic modification protein to a target sequence. In an example embodiment, Mi2/NuRD, Sin3A, or Co-REST recruit HDACs to a target sequence.

[0444] In example embodiments, the epigenetic modification domain can be a eukaryotic or prokaryotic (e.g., bacteria or Archaea) protein. In example embodiments, the eukaryotic protein can be a mammalian, insect, plant, or yeast protein and is not limited to human proteins (e.g., a yeast, insect, plant chromatin modifying protein, such as yeast HATs, HDACs, methyltransferases, etc.

[0445] In one aspect of the invention, is provided a fusion protein (epigenetic modification polypeptide) comprising from N-terminus to C-terminus, an epigenetic modification domain, an XTEN linker, and a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient endonuclease enzyme.

[0446] In aspects, the epigenetic modification polypeptide further comprises a transcriptional activator. In aspects, the transcriptional activator is VP64, p65, RTA, or a combination of two or more thereof. In another aspect, the epigenetic modification polypeptide further comprises one or more nuclear localization sequences. In embodiments, the epigenetic modification polypeptide comprises the nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, the fusion protein comprises the nuclease-deficient DNA endonuclease enzyme.

[0447] In some embodiments, the functional domains associated with the adaptor protein or the CRISPR enzyme is a transcriptional activation domain comprising VP64, p65, MyoD1, HSF1, RTA or SET7/9. Other references herein to activation (or activator) domains in respect of those associated with the adaptor protein(s) include any known transcriptional activation domain and specifically VP64, p65, MyoD1, HSF1, RTA or SET7/9 (see, e.g., US Patent, US11001829B2).

[0448] In certain embodiments, the present invention provides a fusion protein comprising from N-terminus to C-terminus, an RNA-binding sequence, an XTEN linker, and a transcriptional activator. In aspects, the transcriptional activator is VP64, p65, RTA, or a combination of two or more thereof. In aspects, the fusion protein further comprises a demethylation domain, a nuclease-deficient RNA-guided DNA endonuclease enzyme or a nuclease-deficient endonuclease enzyme, a nuclear localization sequence, or a combination of two or more thereof. In embodiments, the fusion protein comprises the nuclease-deficient RNA-guided DNA endonuclease enzyme. In embodiments, the fusion protein comprises the nuclease-deficient DNA endonuclease enzyme.

[0449] Example epigenome modification systems that can be adapted for use with the chimeric IscB systems disclosed herein include US 2020/0003761; WO 2018/053035 ("Targeted DNA Demethylation and Methylation", Jackson Laboratory); WO 2018/148667 ("Reprogramming Cell Aging", Memorial Sloan Kettering Cancer Center); WO 2022/140577 ("Compositions and Methods for Epigenetic Editing", Chroma Medicine); WO 2017/090724 ("DNA Methylation Editing Kit and DNA Methylation Editing Method", Gunma University

NUC); WO 2019/0136229 ("Compositions and Methods of Improving Specificity in Genomic Engineering Using RNA-guided Nucleases", Duke University); WO 2014/059255 (Transcription Activator-like Effector (TALE) – Lysine-specific Demthylase 1 (LSD1) Fusion Proteins", General Hospital Corp.); WO 2014/152432 ("Increasing Specificity for RNA-guided Genome Editing", General Hospital Corp.).

POLYNUCLEOTIDES ENCODING ISCB SYSTEMS AND VECTORS

[0450] The systems herein may comprise one or more polynucleotides. The polynucleotide(s) may comprise coding sequences of components of the systems herein, e.g., chimeric IscB polypeptide(s), ωRNA(s), functional domain(s), donor polynucleotide(s), and/or other components in the systems. The present disclosure further provides vectors or vector systems comprising one or more polynucleotides herein. The vectors or vector systems include those described in the delivery sections herein.

[0451] The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. The term also encompasses nucleic-acid-like structures with synthetic backbones, see, e.g., Eckstein, 1991; Baserga et al., 1992; Milligan, 1993; WO 97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; and Samstag, 1996. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms. A "wild type" can be a base line. As

used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature. The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature. "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions. As used herein, "stringent conditions" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Where reference is made to a polynucleotide sequence, then complementary or partially complementary sequences are also envisaged. These are preferably capable of hybridizing to the reference sequence under highly stringent conditions. "Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may

comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence. As used herein, the term "genomic locus" or "locus" (plural loci) is the specific location of a gene or DNA sequence on a chromosome. A "gene" refers to stretches of DNA or RNA that encode a polypeptide or an RNA chain that has functional role to play in an organism and hence is the molecular unit of heredity in living organisms. For the purpose of this invention, it may be considered that genes include regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions. As used herein, "expression of a genomic locus" or "gene expression" is the process by which information from a gene is used in the synthesis of a functional gene product. The products of gene expression are often proteins, but in non-protein coding genes such as rRNA genes or tRNA genes, the product is functional RNA. The process of gene expression is used by all known life - eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea) and viruses to generate functional products to survive. As used herein "expression" of a gene or nucleic acid encompasses not only cellular gene expression, but also the transcription and translation of nucleic acid(s) in cloning systems and in any other context. As used herein, "expression" also refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond

formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. As used herein, the term "domain" or "protein domain" refers to a part of a protein sequence that may exist and function independently of the rest of the protein chain. As described in aspects of the invention, sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences.

[0452] In an embodiment, the polynucleotide sequence is recombinant DNA. In further embodiments, the polynucleotide sequence further comprises additional sequences as described elsewhere herein. In an embodiment, the nucleic acid sequence is synthesized *in vitro*.

[0453] The present disclosure provides polynucleotide molecules that encode one or more components of the system or chimeric IscB system as referred to in any embodiment herein. In an embodiment, the polynucleotide molecules may comprise further regulatory sequences. By means of guidance and not limitation, the polynucleotide sequence can be part of an expression plasmid, a minicircle, a lentiviral vector, a retroviral vector, an adenoviral or adeno-associated viral vector, a piggyback vector, or a tol2 vector. In an embodiment, the polynucleotide sequence may be a bicistronic expression construct. In further embodiments, the isolated polynucleotide sequence may be incorporated in a cellular genome. In yet further embodiments, the isolated polynucleotide sequence may be part of a cellular genome. In further embodiments, the isolated polynucleotide sequence may be comprised in an artificial chromosome. In an embodiment, the 5' and/or 3' end of the isolated polynucleotide sequence may be modified to improve the stability of the sequence of actively avoid degradation. In an embodiment, the isolated polynucleotide sequence may be comprised in a bacteriophage. In other embodiments, the isolated polynucleotide sequence may be contained in agrobacterium species. In an embodiment, the isolated polynucleotide sequence is lyophilized.

[0454] Aspects of the invention relate to polynucleotide molecules that encode one or more components of one or more systems as described in any of the embodiments herein, wherein at

least one or more regions of the polynucleotide molecule may be codon optimized for expression in an eukaryotic cells. In an embodiment, the polynucleotide molecules that encode one or more components of one or more systems as described in any of the embodiments herein are optimized for expression in a mammalian cell or a plant cell.

[0455] An example of a codon optimized sequence is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e., being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed. In one embodiment, an enzyme coding sequence encoding a DNA/RNA-targeting chimeric IscB system is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In one embodiment, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also

available. In one embodiment, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a chimeric IscB system corresponds to the most frequently used codon for a particular amino acid.

DELIVERY

[0456] The present disclosure also provides delivery systems for introducing components of the systems and compositions herein to cells, tissues, organs, or organisms. A delivery system may comprise one or more delivery vehicles and/or cargos. Exemplary delivery systems and methods include those described in paragraphs [00117] to [00278] of Feng Zhang et al., (WO2016106236A1), and pages 1241-1251 and Table 1 of Lino CA et al., Delivering CRISPR: a review of the challenges and approaches, DRUG DELIVERY, 2018, VOL. 25, NO. 1, 1234–1257, which are incorporated by reference herein in their entireties and can be adapted for use with the IscB proteins disclosed herein.

[0457] In one embodiment, the delivery systems may be used to introduce the components of the systems and compositions to plant cells. For example, the components may be delivered to plant using electroporation, microinjection, aerosol beam injection of plant cell protoplasts, biolistic methods, DNA particle bombardment, and/or Agrobacterium-mediated transformation. Examples of methods and delivery systems for plants include those described in Fu et al., Transgenic Res. 2000 Feb;9(1):11-9; Klein RM, et al., Biotechnology. 1992;24:384-6; Casas AM et al., Proc Natl Acad Sci U S A. 1993 Dec 1; 90(23): 11212–11216; and U.S. Pat. No. 5,563,055, Davey MR et al., Plant Mol Biol. 1989 Sep;13(3):273-85, which are incorporated by reference herein in their entireties.

[0458] The example delivery compositions, systems, and methods described herein related to composition or chimeric IscB systems also apply to functional domains and other components (e.g., other proteins and polynucleotides related to the chimeric IscB systems, such as reverse transcriptase, nucleotide deaminase, retrotransposon, donor polynucleotide, etc.).

Cargos

[0459] The delivery systems may comprise one or more cargos. The cargos may comprise one or more components of the systems and compositions herein. A cargo may comprise one or more of the following: i) a plasmid encoding one or more proteins components in the compositions and systems such as the chimeric IscB ; ii) a plasmid encoding one or more ωRNAs, iii) mRNA of one or more one or more proteins components in the compositions and systems such as the chimeric IscB ; iv) one or more guide RNAs; v) one or more proteins

components in the compositions and systems such as the chimeric IscB polypeptide; vi) any combination thereof.

[0460] In some examples, a cargo may comprise a plasmid encoding one or more proteins components in the compositions and systems such as the chimeric IscB systems disclosed herein. In some cases, the plasmid may also encode a recombination template (e.g., for HDR). In one embodiment, a cargo may comprise mRNA encoding one or more protein components and one or more ωRNA or guide RNAs.

[0461] In some examples, a cargo may comprise one or more protein components and one or more ωRNA or guide RNAs, e.g., in the form of ribonucleoprotein complexes (RNP). The ribonucleoprotein complexes may be delivered by methods and systems herein. In some cases, the ribonucleoprotein may be delivered by way of a polypeptide-based shuttle agent. In one example, the ribonucleoprotein may be delivered using synthetic peptides comprising an endosome leakage domain (ELD) operably linked to a cell penetrating domain (CPD), to a histidine-rich domain and a CPD, e.g., as describe in WO2016161516. RNP may also be used for delivering the compositions and systems to plant cells, e.g., as described in Wu JW, et al., Nat Biotechnol. 2015 Nov;33(11):1162-4.

Physical delivery

[0462] In one embodiment, the cargos may be introduced to cells by physical delivery methods. Examples of physical methods include microinjection, electroporation, and hydrodynamic delivery. Both nucleic acid and proteins may be delivered using such methods. For example, one or more protein components may be prepared *in vitro*, isolated, (refolded, purified if needed), and introduced to cells.

Microinjection

[0463] Microinjection of the cargo directly to cells can achieve high efficiency, e.g., above 90% or about 100%. In one embodiment, microinjection may be performed using a microscope and a needle (e.g., with 0.5–5.0 μm in diameter) to pierce a cell membrane and deliver the cargo directly to a target site within the cell. Microinjection may be used for *in vitro* and *ex vivo* delivery.

[0464] Plasmids comprising coding sequences for one or more protein components and/or ωRNAs, mRNAs, and/or guide RNAs, may be microinjected. In some cases, microinjection may be used i) to deliver DNA directly to a cell nucleus, and/or ii) to deliver mRNA (e.g., *in vitro* transcribed) to a cell nucleus or cytoplasm. In certain examples, microinjection may be

used to delivery ωRNA directly to the nucleus and mRNA to the cytoplasm, e.g., facilitating translation and shuttling of one or more protein components to the nucleus.

[0465] Microinjection may be used to generate genetically modified animals. For example, gene editing cargos may be injected into zygotes to allow for efficient germline modification. Such approach can yield normal embryos and full-term mouse pups harboring the desired modification(s). Microinjection can also be used to provide transiently up- or down- regulate a specific gene within the genome of a cell, e.g., using chimeric IscB system.

Electroporation

[0466] In one embodiment, the cargos and/or delivery vehicles may be delivered by electroporation. Electroporation may use pulsed high-voltage electrical currents to transiently open nanometer-sized pores within the cellular membrane of cells suspended in buffer, allowing for components with hydrodynamic diameters of tens of nanometers to flow into the cell. In some cases, electroporation may be used on various cell types and efficiently transfer cargo into cells. Electroporation may be used for *in vitro* and *ex vivo* delivery.

[0467] Electroporation may also be used to deliver the cargo to into the nuclei of mammalian cells by applying specific voltage and reagents, e.g., by nucleofection. Such approaches include those described in Wu Y, et al. (2015). Cell Res 25:67–79; Ye L, et al. (2014). Proc Natl Acad Sci USA 111:9591–6; Choi PS, Meyerson M. (2014). Nat Commun 5:3728; Wang J, Quake SR. (2014). Proc Natl Acad Sci 111:13157–62. Electroporation may also be used to deliver the cargo *in vivo*, e.g., with methods described in Zuckermann M, et al. (2015). Nat Commun 6:7391.

Hydrodynamic delivery

[0468] Hydrodynamic delivery may also be used for delivering the cargos, e.g., for *in vivo* delivery. In some examples, hydrodynamic delivery may be performed by rapidly pushing a large volume (8–10% body weight) solution containing the gene editing cargo into the bloodstream of a subject (e.g., an animal or human), e.g., for mice, via the tail vein. As blood is incompressible, the large bolus of liquid may result in an increase in hydrodynamic pressure that temporarily enhances permeability into endothelial and parenchymal cells, allowing for cargo not normally capable of crossing a cellular membrane to pass into cells. This approach may be used for delivering naked DNA plasmids and proteins. The delivered cargos may be enriched in liver, kidney, lung, muscle, and/or heart.

Transfection

[0469] The cargos, e.g., nucleic acids, may be introduced to cells by transfection methods for introducing nucleic acids into cells. Examples of transfection methods include calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acid.

Delivery vehicles

[0470] The delivery systems may comprise one or more delivery vehicles. The delivery vehicles may deliver the cargo into cells, tissues, organs, or organisms (e.g., animals or plants). The cargos may be packaged, carried, or otherwise associated with the delivery vehicles. The delivery vehicles may be selected based on the types of cargo to be delivered, and/or the delivery is *in vitro* and/or *in vivo*. Examples of delivery vehicles include vectors, viruses, non-viral vehicles, and other delivery reagents described herein.

[0471] The delivery vehicles in accordance with the present invention may have a greatest dimension (e.g., diameter) of less than 100 microns (μm). In one embodiment, the delivery vehicles have a greatest dimension of less than 10 μm. In one embodiment, the delivery vehicles may have a greatest dimension of less than 2000 nanometers (nm). In one embodiment, the delivery vehicles may have a greatest dimension of less than 1000 nanometers (nm). In one embodiment, the delivery vehicles may have a greatest dimension (e.g., diameter) of less than 900 nm, less than 800 nm, less than 700 nm, less than 600 nm, less than 500 nm, less than 400 nm, less than 300 nm, less than 200 nm, less than 150nm, or less than 100nm, less than 50nm. In one embodiment, the delivery vehicles may have a greatest dimension ranging between 25 nm and 200 nm.

[0472] In one embodiment, the delivery vehicles may be or comprise particles. For example, the delivery vehicle may be or comprise nanoparticles (e.g., particles with a greatest dimension (e.g., diameter) no greater than 1000nm. The particles may be provided in different forms, e.g., as solid particles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of particles, or combinations thereof. Metal, dielectric, and semiconductor particles may be prepared, as well as hybrid structures (e.g., core–shell particles). Nanoparticles may also be used to deliver the compositions and systems to plant cells, e.g., as described in International Patent Publication No. WO 2008042156, US

Publication Application No. US 20130185823, and International Patent Publication No WO 2015/089419.

[0473] In an example embodiment, the delivery vehicles may be or comprise an endogenous LTR retroelement polypeptide or retrotransposons and retroviruses. The endogenous LTR retroelement polypeptide may be an endogenous Gag (Group specific antigen) polypeptide or Gag homolog. A Gag forms a virus-like particle that preferentially binds and facilitates vesicular secretion of its own messenger RNA. In an example embodiment, the Gag homolog is an LTR retrotransposon–derived protein PEG10. See Segel, M.; Lash, B.; Song, J.; Ladha, A.; Liu, C. C.; Jin, X.; Mekhedov, S. L.; Macrae, R. K.; Koonin, E. V.; Zhang, F. Mammalian Retrovirus-like Protein PEG10 Packages Its Own MRNA and Can Be Pseudotyped for MRNA Delivery. Science, 2021, 373, 882–889.

Vectors

[0474] The systems, compositions, and/or delivery systems may comprise one or more vectors. The present disclosure also includes vector systems. A vector system may comprise one or more vectors. In one embodiment, a vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. A vector may be a plasmid, e.g., a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Certain vectors may be capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Some vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. In certain examples, vectors may be expression vectors, e.g., capable of directing the expression of genes to which they are operatively-linked. In some cases, the expression vectors may be for expression in eukaryotic cells. Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0475] Examples of vectors include pGEX, pMAL, pRIT5, E. coli expression vectors (e.g., pTrc, pET 11d, yeast expression vectors (e.g., pYepSec1, pMFa, pJRY88, pYES2, and picZ,

Baculovirus vectors (e.g., for expression in insect cells such as SF9 cells) (e.g., pAc series and the pVL series), mammalian expression vectors (e.g., pCDM8 and pMT2PC.

[0476] A vector may comprise i) one or more protein components encoding sequence(s), and/or ii) a single, or at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 14, at least 16, at least 32, at least 48, at least 50 guide RNA(s) encoding sequences. In a single vector there can be a promoter for each RNA coding sequence. Alternatively or additionally, in a single vector, there may be a promoter controlling (e.g., driving transcription and/or expression) multiple RNA encoding sequences.

[0477] Furthermore, that compositions or systems may be delivered via a vector, e.g., a separate vector or the same vector that is encoding the complex. When provided by a separate vector, the RNA that targets chimeric IscB system expression can be administered sequentially or simultaneously. When administered sequentially, the RNA that targets chimeric IscB system expression is to be delivered after the RNA that is intended for e.g., gene editing or gene engineering. This period may be a period of minutes (e.g., 5 minutes, 10 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes). This period may be a period of hours (e.g., 2 hours, 4 hours, 6 hours, 8 hours, 12 hours, 24 hours). This period may be a period of days (e.g., 2 days, 3 days, 4 days, 7 days). This period may be a period of weeks (e.g., 2 weeks, 3 weeks, 4 weeks). This period may be a period of months (e.g., 2 months, 4 months, 8 months, 12 months). This period may be a period of years (2 years, 3 years, 4 years). In this fashion, the chimeric IscB system associates with a first ωRNA molecule capable of hybridizing to a first target, such as a genomic locus or loci of interest and undertakes the function(s) desired of the system (e.g., gene engineering); and subsequently the chimeric IscB systems may then associate with the second ωRNA molecule capable of hybridizing to the sequence comprising at least part of the chimeric IscB systems. Where the guide RNA targets the sequences encoding expression of the chimeric IscB systems, the enzyme becomes impeded, and the system becomes self-inactivating. In the same manner, RNA that targets chimeric IscB system nuclease expression applied via, for example liposome, lipofection, particles, micro-vesicles as explained herein, may be administered sequentially or simultaneously. Similarly, self-inactivation may be used for inactivation of one or more guide RNA used to target one or more targets.

Regulatory elements

[0478] A vector may comprise one or more regulatory elements. The regulatory element(s) may be operably linked to coding sequences of chimeric IscB systems, accessory proteins,

ωRNA scaffold and/or guide RNA or combination thereof. The term "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). In certain examples, a vector may comprise: a first regulatory element operably linked to a nucleotide sequence encoding a chimeric IscB system, and a second regulatory element operably linked to a nucleotide sequence encoding a ωRNA or guide RNA.

[0479] Examples of regulatory elements include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific.

[0480] Examples of promoters include one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter.

Viral vectors

[0481] The cargos may be delivered by viruses. In one embodiment, viral vectors are used. A viral vector may comprise virally-derived DNA or RNA sequences for packaging into a virus

(e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Viruses and viral vectors may be used for *in vitro*, *ex vivo*, and/or *in vivo* deliveries.

Adeno associated virus (AAV)

[0482] The systems and compositions herein may be delivered by adeno associated virus (AAV). AAV vectors may be used for such delivery. AAV, of the Dependovirus genus and Parvoviridae family, is a single stranded DNA virus. In one embodiment, AAV may provide a persistent source of the provided DNA, as AAV delivered genomic material can exist indefinitely in cells, e.g., either as exogenous DNA or, with some modification, be directly integrated into the host DNA. In one embodiment, AAV do not cause or relate with any diseases in humans. The virus itself is able to efficiently infect cells while provoking little to no innate or adaptive immune response or associated toxicity.

[0483] Examples of AAV that can be used herein include AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, and AAV-9. The type of AAV may be selected with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. AAV-2-based vectors were originally proposed for CFTR delivery to CF airways, other serotypes such as AAV-1, AAV-5, AAV-6, and AAV-9 exhibit improved gene transfer efficiency in a variety of models of the lung epithelium. Examples of cell types targeted by AAV are described in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)), and shown in Table 5 as follows:

Table 5. Examples of cell types targeted by AAV.

Cell Line	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND

A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
Monocytes	1111	100	ND	ND	125	1429	ND	ND
Immature DC	2500	100	ND	ND	222	2857	ND	ND
Mature DC	2222	100	ND	ND	333	3333	ND	ND

[0484] The AAV particles may be created in HEK 293 T cells. Once particles with specific tropism have been created, they are used to infect the target cell line much in the same way that native viral particles do. This may allow for persistent presence of the components in the infected cell type, and what makes this version of delivery particularly suited to cases where long-term expression is desirable. Examples of doses and formulations for AAV that can be used include those describe in US Patent Nos. 8,454,972 and 8,404,658.

[0485] Various strategies may be used for delivery the systems and compositions herein with AAVs. In some examples, coding sequences of chimeric IscB system and ωRNA may be packaged directly onto one DNA plasmid vector and delivered via one AAV particle. In some examples, AAVs may be used to deliver ωRNAs into cells that have been previously engineered to express chimeric IscB systems. In some examples, coding sequences of chimeric IscB systems and ωRNA may be made into two separate AAV particles, which are used for co-transfection of target cells. In some examples, markers, tags, and other sequences may be packaged in the same AAV particles as coding sequences of chimeric IscB systems and/or ωRNAs.

Lentiviruses

[0486] The systems and compositions herein may be delivered by lentiviruses. Lentiviral vectors may be used for such delivery. Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells.

[0487] Examples of lentiviruses include human immunodeficiency virus (HIV), which may use its envelope glycoproteins of other viruses to target a broad range of cell types; minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV), which may be used for ocular therapies. In an embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the nucleic acid-targeting system herein.

[0488] Lentiviruses may be pseudo-typed with other viral proteins, such as the G protein of vesicular stomatitis virus. In doing so, the cellular tropism of the lentiviruses can be altered to be as broad or narrow as desired. In some cases, to improve safety, second- and third-generation lentiviral systems may split essential genes across three plasmids, which may reduce the likelihood of accidental reconstitution of viable viral particles within cells.

[0489] In some examples, leveraging the integration ability, lentiviruses may be used to create libraries of cells comprising various genetic modifications, e.g., for screening and/or studying genes and signaling pathways.

Adenoviruses

[0490] The systems and compositions herein may be delivered by adenoviruses. Adenoviral vectors may be used for such delivery. Adenoviruses include nonenveloped viruses with an icosahedral nucleocapsid containing a double stranded DNA genome. Adenoviruses may infect dividing and non-dividing cells. In one embodiment, adenoviruses do not integrate into the genome of host cells, which may be used for limiting off-target effects of systems in gene editing applications.

Viral vehicles for delivery to plants

[0491] The systems and compositions may be delivered to plant cells using viral vehicles. In one embodiment, the compositions and systems may be introduced in the plant cells using a plant viral vector (e.g., as described in Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323). Such viral vector may be a vector from a DNA virus, e.g., geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). The viral vector may be a vector from an RNA virus, e.g., tobravirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses may be non-integrative vectors.

Non-viral vehicles

[0492] The delivery vehicles may comprise non-viral vehicles. In general, methods and vehicles capable of delivering nucleic acids and/or proteins may be used for delivering the systems compositions herein. Examples of non-viral vehicles include lipid nanoparticles, cell-penetrating peptides (CPPs), DNA nanoclews, gold nanoparticles, streptolysin O,

multifunctional envelope-type nanodevices (MENDs), lipid-coated mesoporous silica particles, and other inorganic nanoparticles.

Lipid particles

[0493] The delivery vehicles may comprise lipid particles, e.g., lipid nanoparticles (LNPs) and liposomes.

Lipid nanoparticles (LNPs)

[0494] LNPs may encapsulate nucleic acids within cationic lipid particles (e.g., liposomes), and may be delivered to cells with relative ease. In some examples, lipid nanoparticles do not contain any viral components, which helps minimize safety and immunogenicity concerns. Lipid particles may be used for *in vitro*, *ex vivo*, and *in vivo* deliveries. Lipid particles may be used for various scales of cell populations.

[0495] In some examples. LNPs may be used for delivering DNA molecules (e.g., those comprising coding sequences of chimeric IscB system and/or ωRNA) and/or RNA molecules (e.g., mRNA of chimeric IscB systems, ωRNA or gRNAs). In certain cases, LNPs may be use for delivering RNP complexes of chimeric IscB system/ωRNA.

[0496] Components in LNPs may comprise cationic lipids 1,2- dilineoyl-3-dimethylammonium-propane (DLinDAP), l,2-dilinoleyloxy-3-N,N- dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), l,2-dilinoleyl-4-(2-dimethylaminoethyl)-[l,3]-dioxolane (DLinKC2-DMA), (3- o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-l,2-dimyristoyl-sn-glycol (PEG-S-DMG), R-3-[(ro-methoxy-poly(ethylene glycol)2000) carbamoyl]-l,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG, and any combination thereof. Preparation of LNPs and encapsulation may be adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011).

Liposomes

[0497] In one embodiment, a lipid particle may be liposome. Liposomes are spherical vesicle structures composed of a uni- or multi-lamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. In one embodiment, liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB).

[0498] Liposomes can be made from several different types of lipids, e.g., phospholipids. A liposome may comprise natural phospholipids and lipids such as l,2-distearoryl-sn-glycero-

3 -phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines, monosialoganglioside, or any combination thereof.

[0499] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, liposomes may further comprise cholesterol, sphingomyelin, and/or l,2-dioleoyl-sn-glycero-3- phosphoethanolamine (DOPE), e.g., to increase stability and/or to prevent the leakage of the liposomal inner cargo.

Stable nucleic-acid-lipid particles (SNALPs)

[0500] In one embodiment, the lipid particles may be stable nucleic acid lipid particles (SNALPs). SNALPs may comprise an ionizable lipid (DLinDMA) (e.g., cationic at low pH), a neutral helper lipid, cholesterol, a diffusible polyethylene glycol (PEG)-lipid, or any combination thereof. In some examples, SNALPs may comprise synthetic cholesterol, dipalmitoylphosphatidylcholine, 3-N-[(w-methoxy polyethylene glycol)2000)carbamoyl]-l,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane. In some examples, SNALPs may comprise synthetic cholesterol, 1,2-distearoyl-sn-glycero-3-phosphocholine, PEG- cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA).

Other lipids

[0501] The lipid particles may also comprise one or more other types of lipids, e.g., cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[l,3]- dioxolane (DLin-KC2-DMA), DLin-KC2-DMA4, C12- 200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG.

Lipoplexes/polyplexes

[0502] In one embodiment, the delivery vehicles comprise lipoplexes and/or polyplexes. Lipoplexes may bind to negatively charged cell membrane and induce endocytosis into the cells. Examples of lipoplexes may be complexes comprising lipid(s) and non-lipid components. Examples of lipoplexes and polyplexes include FuGENE-6 reagent, a non-liposomal solution containing lipids and other components, zwitterionic amino lipids (ZALs), Ca2þ (e.g., forming DNA/Ca^{2+} micro-complexes), polyethenimine (PEI) (e.g., branched PEI), and poly(L-lysine) (PLL).

Cell penetrating peptides

[0503] In one embodiment, the delivery vehicles comprise cell penetrating peptides (CPPs). CPPs are short peptides that facilitate cellular uptake of various molecular cargo (e.g., from nanosized particles to small chemical molecules and large fragments of DNA).

[0504] CPPs may be of different sizes, amino acid sequences, and charges. In some examples, CPPs can translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPPs may be introduced into cells via different mechanisms, e.g., direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure.

[0505] CPPs may have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. Another type of CPPs is the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1). Examples of CPPs include to Penetratin, Tat (48-60), Transportan, and (R-AhX-R4) (Ahx refers to aminohexanoyl), Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin β3 signal peptide sequence, polyarginine peptide Args sequence, Guanine rich-molecular transporters, and sweet arrow peptide. Examples of CPPs and related applications also include those described in US Patent No. 8,372,951.

[0506] CPPs can be used for *in vitro* and *ex vivo* work quite readily, and extensive optimization for each cargo and cell type is usually required. In some examples, CPPs may be covalently attached to the chimeric IscB system directly, which is then complexed with the ωRNA and delivered to cells. In some examples, separate delivery of CPP–IscB and CPP–ωRNA to multiple cells may be performed. CPP may also be used to delivery RNPs.

[0507] CPPs may be used to deliver the compositions and systems to plants. In some examples, CPPs may be used to deliver the components to plant protoplasts, which are then regenerated to plant cells and further to plants.

DNA nanoclews

[0508] In one embodiment, the delivery vehicles comprise DNA nanoclews. A DNA nanoclew refers to a sphere-like structure of DNA (e.g., with a shape of a ball of yarn). The

nanoclew may be synthesized by rolling circle amplification with palindromic sequences that aide in the self-assembly of the structure. The sphere may then be loaded with a payload. An example of DNA nanoclew is described in Sun W et al, J Am Chem Soc. 2014 Oct 22;136(42):14722-5; and Sun W et al, Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. DNA nanoclew may have a palindromic sequence to be partially complementary to the guide RNA within the chimeric IscB system: ωRNA ribonucleoprotein complex. A DNA nanoclew may be coated, e.g., coated with PEI to induce endosomal escape.

Gold nanoparticles

[0509] In one embodiment, the delivery vehicles comprise gold nanoparticles (also referred to AuNPs or colloidal gold). Gold nanoparticles may form complex with cargos, e.g., chimeric IscB system: ωRNA RNP. Gold nanoparticles may be coated, e.g., coated in a silicate and an endosomal disruptive polymer, PAsp(DET). Examples of gold nanoparticles include AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, and those described in Mout R, et al. (2017). ACS Nano 11:2452–8; Lee K, et al. (2017). Nat Biomed Eng 1:889–901.

iTOP

[0510] In one embodiment, the delivery vehicles comprise iTOP. iTOP refers to a combination of small molecules drives the highly efficient intracellular delivery of native proteins, independent of any transduction peptide. iTOP may be used for induced transduction by osmocytosis and propanebetaine, using NaCl-mediated hyperosmolality together with a transduction compound (propanebetaine) to trigger macropinocytotic uptake into cells of extracellular macromolecules. Examples of iTOP methods and reagents include those described in D'Astolfo DS, Pagliero RJ, Pras A, et al. (2015). Cell 161:674–690.

Polymer-based particles

[0511] In one embodiment, the delivery vehicles may comprise polymer-based particles (e.g., nanoparticles). In one embodiment, the polymer-based particles may mimic a viral mechanism of membrane fusion. The polymer-based particles may be a synthetic copy of Influenza virus machinery and form transfection complexes with various types of nucleic acids ((siRNA, miRNA, plasmid DNA or shRNA, mRNA) that cells take up via the endocytosis pathway, a process that involves the formation of an acidic compartment. The low pH in late endosomes acts as a chemical switch that renders the particle surface hydrophobic and facilitates membrane crossing. Once in the cytosol, the particle releases its payload for cellular action. This Active Endosome Escape technology is safe and maximizes transfection efficiency

as it is using a natural uptake pathway. In one embodiment, the polymer-based particles may comprise alkylated and carboxyalkylated branched polyethyleneimine. In some examples, the polymer-based particles are VIROMER, e.g., VIROMER RNAi, VIROMER RED, VIROMER mRNA. Example methods of delivering the systems and compositions herein include those described in Bawage SS et al., Synthetic mRNA expressed Cas13a mitigates RNA virus infections, www.biorxiv.org/content/10.1101/370460v1.full doi: doi.org/10.1101/370460, Viromer® RED, a powerful tool for transfection of keratinocytes. doi: 10.13140/RG.2.2.16993.61281, Viromer® Transfection - Factbook 2018: technology, product overview, users' data., doi:10.13140/RG.2.2.23912.16642.

Streptolysin O (SLO)

[0512] The delivery vehicles may be streptolysin O (SLO). SLO is a toxin produced by Group A streptococci that works by creating pores in mammalian cell membranes. SLO may act in a reversible manner, which allows for the delivery of proteins (e.g., up to 100 kDa) to the cytosol of cells without compromising overall viability. Examples of SLO include those described in Sierig G, et al. (2003). Infect Immun 71:446–55; Walev I, et al. (2001). Proc Natl Acad Sci U S A 98:3185–90; Teng KW, et al. (2017). Elife 6:e25460.

Multifunctional envelope-type nanodevice (MEND)

[0513] The delivery vehicles may comprise multifunctional envelope-type nanodevice (MENDs). MENDs may comprise condensed plasmid DNA, a PLL core, and a lipid film shell. A MEND may further comprise cell-penetrating peptide (e.g., stearyl octaarginine). The cell penetrating peptide may be in the lipid shell. The lipid envelope may be modified with one or more functional components, e.g., one or more of: polyethylene glycol (e.g., to increase vascular circulation time), ligands for targeting of specific tissues/cells, additional cell-penetrating peptides (e.g., for greater cellular delivery), lipids to enhance endosomal escape, and nuclear delivery tags. In some examples, the MEND may be a tetra-lamellar MEND (T-MEND), which may target the cellular nucleus and mitochondria. In certain examples, a MEND may be a PEG-peptide-DOPE-conjugated MEND (PPD-MEND), which may target bladder cancer cells. Examples of MENDs include those described in Kogure K, et al. (2004). J Control Release 98:317–23; Nakamura T, et al. (2012). Acc Chem Res 45:1113–21.

Lipid-coated mesoporous silica particles

[0514] The delivery vehicles may comprise lipid-coated mesoporous silica particles. Lipid-coated mesoporous silica particles may comprise a mesoporous silica nanoparticle core and a

lipid membrane shell. The silica core may have a large internal surface area, leading to high cargo loading capacities. In one embodiment, pore sizes, pore chemistry, and overall particle sizes may be modified for loading different types of cargos. The lipid coating of the particle may also be modified to maximize cargo loading, increase circulation times, and provide precise targeting and cargo release. Examples of lipid-coated mesoporous silica particles include those described in Du X, et al. (2014). Biomaterials 35:5580–90; Durfee PN, et al. (2016). ACS Nano 10:8325–45.

Inorganic nanoparticles

[0515] The delivery vehicles may comprise inorganic nanoparticles. Examples of inorganic nanoparticles include carbon nanotubes (CNTs) (e.g., as described in Bates K and Kostarelos K. (2013). Adv Drug Deliv Rev 65:2023–33.), bare mesoporous silica nanoparticles (MSNPs) (e.g., as described in Luo GF, et al. (2014). Sci Rep 4:6064), and dense silica nanoparticles (SiNPs) (as described in Luo D and Saltzman WM. (2000). Nat Biotechnol 18:893–5).

Exosomes

[0516] The delivery vehicles may comprise exosomes. Exosomes include membrane bound extracellular vesicles, which can be used to contain and delivery various types of biomolecules, such as proteins, carbohydrates, lipids, and nucleic acids, and complexes thereof (e.g., RNPs). Examples of exosomes include those described in Schroeder A, et al., J Intern Med. 2010 Jan;267(1):9-21; El-Andaloussi S, et al., Nat Protoc. 2012 Dec;7(12):2112-26; Uno Y, et al., Hum Gene Ther. 2011 Jun;22(6):711-9; Zou W, et al., Hum Gene Ther. 2011 Apr;22(4):465-75.

[0517] In some examples, the exosome may form a complex (e.g., by binding directly or indirectly) to one or more components of the cargo. In certain examples, a molecule of an exosome may be fused with first adapter protein and a component of the cargo may be fused with a second adapter protein. The first and the second adapter protein may specifically bind each other, thus associating the cargo with the exosome. Examples of such exosomes include those described in Ye Y, et al., Biomater Sci. 2020 Apr 28. doi: 10.1039/d0bm00427h.

Retrovirus Like Delivery Systems

[0518] The delivery vehicle may comprise a retro-virus like protein, such as PEG10, which is capable of incorporating a cargo into a virus-like particle. As such systems can be re-programmed to package specific cargos, polynucleotides encoding components of the IscB systems disclosed herein may be further modified with a recognition sequence that leads to

selective packaging of the IscB components into such retro-virus like VLPs. Said VLPs may be further modified with fusogenic proteins that impart tissue or cell specificity. Example systems are disclosed in Segal et al. Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudo-typed for mRNA delivery. 373 Science, 882-889 (2021), which is incorporated herein by reference. Engineered VLPs, including DNA-free virus-like particles have been developed that can address cargo packaging, release and localization issues and may be engineered for specific cellular tropism, including particles engineered and optimized for delivery of base editors and ribonucleoproteins. See, e.g. Baskota *et al.*, *Cell.* 2022 Jan 20;185(2):250-265.e16; doi: 10.1016/j.cell.2021.12.021, incorporated herein by reference (development of retroviral-based engineered VLP platform for packaging and delivering base editors and Cas9 nuclease, with potential advantages of both viral and nonviral delivery strategies).

GENETICALLY MODIFIED CELLS AND ORGANISMS

[0519] The present disclosure further provides cells comprising one or more components of the chimeric IscB compositions and systems disclosed herein. Also provided include cells modified by the systems and methods herein, and cell cultures, tissues, organs, organism comprising such cells or progeny thereof. In one embodiment, the present disclosure provides a method of modifying a cell or organism. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The mammalian cell many be a non-human primate, bovine, porcine, rodent or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry, fish or shrimp. The cell may be a therapeutic T cell or antibody-producing B-cell. The cell may also be a plant cell. The plant cell may be of a crop plant such as cassava, corn, sorghum, wheat, or rice. The plant cell may also be of an algae, tree or vegetable. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol or other desired cellular output. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell include an alteration that changes the biologic product produced.

[0520] In one embodiment, one or more polynucleotide molecules, vectors, or vector systems driving expression of one or more elements of the compositions, systems, or delivery systems comprising one or more elements of the nucleic acid-targeting system are introduced into a host cell such that expression of the elements of the nucleic acid-targeting system direct

formation of a nucleic acid-targeting complex at one or more target sites. In an embodiment of the invention the host cell may be a eukaryotic cell, a prokaryotic cell, or a plant cell.

[0521] In one embodiment, the host cell is a cell of a cell line. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In one embodiment, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In one embodiment, a cell transiently transfected with the components of a system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In one embodiment, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0522] Further intended are isolated human cells or tissues, plants or non-human animals comprising one or more of the polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein. In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including (isolated) stem cells, and progeny thereof.

[0523] In an embodiment, the plants or non-human animals comprise at least one of the system components, polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein at least one tissue type of the plant or non-human animal. In an embodiment, non-human animals comprise at least one of the system components, polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein in at least one tissue type. In an embodiment, the presence of the system components is transient, in that they are degraded over time. In an embodiment, expression of the components of the systems and compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is limited to certain tissue types or regions in the plant or non-human animal. In an embodiment, the expression of the components of the systems and compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent of a physiological cue. In an embodiment, expression of the components of the systems and compositions described in any of the embodiments comprised in polynucleotide molecules,

vectors, vector systems, or cells may be triggered by an exogenous molecule. In an embodiment, expression of the components of the systems and compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent on the expression of a non-cas molecule in the plant or non-human animal.

APPLICATIONS AND USES IN GENERAL

[0524] The systems, the vector systems, the vectors and the compositions described herein may be used in various nucleic acids-targeting applications, altering or modifying synthesis of a gene product, such as a protein, nucleic acids cleavage, nucleic acids editing, nucleic acid splicing; trafficking of target nucleic acids, tracing of target nucleic acids, isolation of target nucleic acids, visualization of target nucleic acids, etc.

[0525] Aspects of the invention thus also encompass methods and uses of the compositions and systems described herein in genome engineering, e.g., for altering or manipulating the expression of one or more genes or the one or more gene products, in prokaryotic or eukaryotic cells, *in vitro*, *in vivo* or *ex vivo*. In some examples, the target polynucleotides are target sequences within genomic DNA, including nuclear genomic DNA, mitochondrial DNA, or chloroplast DNA.

[0526] Typically, in the context of a nucleic acid-targeting system, formation of a nucleic acid-targeting complex (comprising a ωRNA or guide RNA hybridized to a target sequence and complexed with one or more nucleic acid-targeting effector proteins) results in cleavage of one or both DNA or RNA strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest).

[0527] In one embodiment, the present disclosure provides a method of targeting a polynucleotide, comprising contacting a sample (such as cell, population of cells, tissue, organ, or an organism) that comprises a target polynucleotide with the composition, systems, polynucleotide(s), or vector(s). The contacting may result in modification of a gene product or modification of the amount or expression of a gene product. In some examples, the target sequence of the polynucleotide is a disease-associated target sequence.

[0528] In one embodiment, the present disclosure provides a method of modifying target polynucleotides comprising delivering the composition, the one or more polynucleotides of 2,

or one or more vectors to a cell or population of cells comprising the target polynucleotides, wherein the complex directs the reverse transcriptase to the target sequence and the reverse transcriptase facilitates insertion of the donor sequence from the ωRNA into the target polynucleotide.

[0529] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown and may be at a normal or abnormal level.

[0530] The target polynucleotide of a complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a TAM (target adjacent motif); that is, a short sequence recognized by the complex. The precise sequence and length requirements for the TAM differ depending on the chimeric IscB polypeptide used, but TAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) A skilled person will be able to identify further TAM sequences for use with a given chimeric IscB polypeptide. As discussed in further detail above, engineering of the TAM Interacting domain may allow programing of TAM specificity, improve target site recognition fidelity, and increase the versatility of the chimeric IscB system, genome engineering platform. Chimeric IscB systems may be engineered to alter their TAM specificity.

[0531] In an embodiment the IscB TAM is ATNA where N is any nucleotide. In an embodiment, the IscB TAM is ATGA, ATAA, ATAAA, or ATN. In one embodiment, the IscB is derived from *Ignatius tetrasporus* and the TAM is NNG.

[0532] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown and may be at a normal or abnormal level.

[0533] Aspects of the invention relate to a method of targeting a polynucleotide, comprising contacting a sample that comprises the polynucleotide with a chimeric IscB composition, or vector encoding a chimeric IscB system or a delivery composition comprising the chimeric IscB system, as described in any embodiment herein. In an embodiment, a target polynucleotide is contacted with at least two different chimeric IscB systems. The two different chimeric IscB systems may have different target polynucleotide specificities, or degrees of specificity. In an embodiment, the two different chimeric IscB systems have a different TAM specificity.

[0534] Also envisaged are methods of targeting a polynucleotide, comprising contacting a sample that comprises the polynucleotide with the composition, vectors, polynucleotides, and delivery systems disclosed herein, wherein contacting results in modification of a gene product or modification of the amount or expression of a gene product. In an embodiment, the expression of the targeted gene product is increased by the method. In an embodiment, the expression of the targeted gene product is increased by at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, p at least 90%,

at least 95%, 100%. In an embodiment, the expression of the targeted gene product is increased at least 1.5-fold, at least 2-fold, at least 2.5-fold, at least 3-fold, at least 3.5-fold, at least 3.5-fold, at least 4-fold, at least 4.5-fold, at least 5-fold, at least 10-fold, at least 10-fold, at least 15-fold, at least 20-fold, at least 25-fold, at least 50-fold, at least 100-fold. In an embodiment, the expression of the targeted gene product is reduced by at least 10%, by at least 15%, by at least 20%, by at least 25%, by at least 30%, by at least 35%, by at least 40%, by at least 45%, by at least 50%, by at least 55%, by at least 60%, by at least 65%, by at least 70%, by at least 75%, by at least 80%, by at least 85%, by at least 90%, by at least 95%, by at least 100%. In an embodiment, the expression of the targeted gene product is reduced at least 1.5-fold, at least 2-fold, at least 2.5-fold, at least 3-fold, at least 3.5-fold, at least 3.5-fold, at least 4-fold, at least 4.5-fold, at least 5-fold, at least 10-fold, at least 10-fold, at least 15-fold, at least 20-fold, at least 25-fold, at least 50-fold, at least 100-fold. In alternative embodiments, the expression of the targeted gene product is reduced by the method. In further embodiments, expression of the targeted gene may be completely eliminated, or may be considered eliminated as remnant expression levels of the targeted gene fall below the detection limit of methods known in the art that are used to quantify, detect, or monitor expression levels of genes.

[0535] In one embodiment, one or more polynucleotide molecules, vectors, or vector systems driving expression of one or more elements comprising one or more elements of the chimeric IscB system are introduced into a host cell such that expression of the elements of the chimeric IscB system direct formation of a nucleic acid-targeting complex at one or more target sites. In an embodiment of the invention the host cell may be a eukaryotic cell, a prokaryotic cell, or a plant cell.

[0536] In one embodiment, the host cell is a cell of a cell line. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In one embodiment, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In one embodiment, a cell transiently transfected with the components of a composition or system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In one embodiment, cells transiently or non-transiently transfected with

one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0537] Further intended are isolated human cells or tissues, plants or non-human animals comprising one or more of the polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein. In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including (isolated) stem cells, and progeny thereof.

[0538] In an embodiment, the plants or non-human animals comprise at least one of the compositions, polynucleotide molecules, vectors, vector systems, delivery systems or cells described in any of the embodiments herein at least one tissue type of the plant or non-human animal. In certain embodiment, non-human animals comprise at least one of the compositions, polynucleotide molecules, vectors, vector systems, delivery systems, or cells described in any of the embodiments herein in at least one tissue type. In an embodiment, the presence of the compositions is transient, in that they are degraded over time. In an embodiment, expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is limited to certain tissue types or regions in the plant or non-human animal. In an embodiment, the expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent of a physiological cue. In an embodiment, expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells may be triggered by an exogenous molecule. In an embodiment, expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent on the expression of a non-Cas molecule in the plant or non-human animal.

[0539] In one aspect, the invention provides methods for using one or more elements of a nucleic acid-targeting system. The nucleic acid-targeting complex of the invention provides an effective means for modifying a target DNA or RNA (single or double stranded, linear or super-coiled). The nucleic acid-targeting complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target DNA or RNA in a multiplicity of cell types. As such, the nucleic acid-targeting complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary nucleic acid-targeting complex comprises a DNA or

RNA-targeting effector protein complexed with a ωRNA or guide RNA hybridized to a target sequence within the target locus of interest.

[0540] In one embodiment, this invention provides a method of cleaving a target polynucleotide. The method may comprise modifying a target polynucleotide using a nucleic acid-targeting complex that binds to the target polynucleotide and effect cleavage of said target polynucleotide. In an embodiment, the nucleic acid-targeting complex of the invention, when introduced into a cell, may create a break (e.g., a single or a double strand break) in the polynucleotide sequence. For example, the method can be used to cleave a disease polynucleotide in a cell. For example, an exogenous template comprising a sequence to be integrated flanked by an upstream sequence and a downstream sequence may be introduced into a cell. The upstream and downstream sequences share sequence similarity with either side of the site of integration in the polynucleotide. The exogenous template comprises a sequence to be integrated (e.g., a mutated RNA). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotide encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function. The upstream and downstream sequences in the recombination template are selected to promote recombination between the RNA sequence of interest and the recombination. The upstream sequence is a polynucleotide sequence that shares sequence similarity with the sequence upstream of the targeted site for integration. Similarly, the downstream sequence is a polynucleotide sequence that shares sequence similarity with the polynucleotide sequence downstream of the targeted site of integration. The upstream and downstream sequences in the recombination template can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted sequence. Preferably, the upstream and downstream sequences in the recombination template have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted sequence. In some methods, the upstream and downstream sequences in the recombination template have about 99% or 100% sequence identity with the targeted sequence. An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about

600 bp to about 1000 bp, or more particularly about 700 bp to about 1000 bp. In some methods, the recombination template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The recombination template of the invention can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996). In a method for modifying a target sequence by integrating an recombination template, a break (e.g., double or single stranded break in double or single stranded DNA or RNA) is introduced into the DNA or RNA sequence by the nucleic acid-targeting complex, the break is repaired via homologous recombination with an recombination template such that the template is integrated into the target. The presence of a double-stranded break facilitates integration of the template. In other embodiments, this invention provides a method of modifying expression of a RNA in a eukaryotic cell. The method comprises increasing or decreasing expression of a target polynucleotide by using a nucleic acid-targeting complex that binds to the DNA or RNA (e.g., mRNA or pre-mRNA). In some methods, a target can be inactivated to affect the modification of the expression in a cell. For example, upon the binding of a nucleic acid-targeting complex to a target sequence in a cell, the target is inactivated such that the sequence is not translated, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein or microRNA or pre-microRNA transcript is not produced. The target of a nucleic acid-targeting complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., ncRNA, lncRNA, tRNA, or rRNA). Examples of target RNA include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated polynucleotide. Examples of target polynucleotide include a disease associated polynucleotide. A “disease-associated” polynucleotide refers to any polynucleotide which is yielding translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated polynucleotide also

refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The translated products may be known or unknown and may be at a normal or abnormal level. The target RNA of a nucleic acid-targeting complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target RNA can be a RNA residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., ncRNA, lncRNA, tRNA, or rRNA).

[0541] In one embodiment, the method may comprise allowing a composition to bind to the target DNA or RNA to effect cleavage of said target DNA or RNA thereby modifying the target DNA or RNA, wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA hybridized to a target sequence within said target DNA or RNA. In one aspect, the invention provides a method of modifying expression of DNA or RNA in a eukaryotic cell. In one embodiment, the method comprises allowing a nucleic acid-targeting complex to bind to the DNA or RNA such that said binding results in increased or decreased expression of said DNA or RNA; wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a ωRNA or guide RNA. Similar considerations and conditions apply as above for methods of modifying a target DNA or RNA. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention. In one aspect, the invention provides for methods of modifying a target DNA or RNA in a eukaryotic cell, which may be *in vivo*, *ex vivo* or *in vitro*. In one embodiment, the method comprises sampling a cell or population of cells from a human or non-human animal and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells. The compositions as described in any embodiment herein may be used to detect nucleic acid identifiers. Nucleic acid identifiers are non-coding nucleic acids that may be used to identify a particular article. Example nucleic acid identifiers, such as DNA watermarks, are described in Heider and Barnekow. "DNA watermarks: A proof of concept" BMC Molecular Biology 9:40 (2008). The nucleic acid identifiers may also be a nucleic acid barcode. A nucleic acid-based barcode is a short sequence of nucleotides (for example, DNA, RNA, or combinations thereof) that is used as an identifier for an associated molecule, such as a target molecule and/or target nucleic acid. A nucleic acid barcode can have a length of at least, for example, 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, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides, and can be in single- or double-stranded form. One or more nucleic acid barcodes can be attached, or "tagged," to a target molecule and/or target nucleic acid. This attachment can be direct (for example, covalent or non-covalent binding of the barcode to the target molecule) or indirect (for example, via an additional molecule, for example, a specific binding agent, such as an antibody (or other protein) or a barcode receiving adaptor (or other nucleic acid molecule). Target molecule and/or target nucleic acids can be labeled with multiple nucleic acid barcodes in combinatorial fashion, such as a nucleic acid barcode concatemer. Typically, a nucleic acid barcode is used to identify target molecules and/or target nucleic acids as being from a particular compartment (for example a discrete volume), having a particular physical property (for example, affinity, length, sequence, etc.), or having been subject to certain treatment conditions. Target molecule and/or target nucleic acid can be associated with multiple nucleic acid barcodes to provide information about all of these features (and more). Methods of generating nucleic acid- barcodes are disclosed, for example, in International Patent Application Publication No. WO/2014/047561.

[0542] In an embodiment, compositions induce a double strand break for the purpose of inducing HDR-mediated correction. In a further embodiment, two or more guide RNAs complexing with chimeric IscB system or an ortholog or homolog thereof, may be used to induce multiplexed breaks for purpose of inducing HDR-mediated correction.

[0543] A recombination template nucleic acid, as that term is used herein, refers to a nucleic acid sequence which can be used in conjunction with compositions discloser herein to alter the structure of a target position. In an embodiment, the target nucleic acid is modified to have some or all of the sequence of the recombination template nucleic acid, typically at or near cleavage site(s). In an embodiment, the recombination template nucleic acid is single stranded. In an alternate embodiment, the recombination template nucleic acid is double stranded. In an embodiment, the recombination template nucleic acid is DNA, e.g., double stranded DNA. In an alternate embodiment, the recombination template nucleic acid is single stranded DNA.

[0544] In one embodiment, a recombination template is provided to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a nucleic acid-targeting effector protein as a part of a nucleic acid-targeting complex.

[0545] A recombination template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. A recombination template polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In one embodiment, the recombination template polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a recombination template polynucleotide might overlap with one or more nucleotides of a target sequences (e.g., about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more nucleotides). In one embodiment, when a recombination template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the recombination template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[0546] In an embodiment, the recombination template nucleic acid alters the structure of the target position by participating in homologous recombination. In an embodiment, the recombination template nucleic acid alters the sequence of the target position. In an embodiment, the recombination template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

[0547] The recombination template sequence may undergo a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the recombination template nucleic acid may include sequence that corresponds to a site on the target sequence that is cleaved by an chimeric IscB system mediated cleavage event. In an embodiment, the recombination template nucleic acid may include sequence that corresponds to both, a first site on the target sequence that is cleaved in a first chimeric IscB system mediated event and a second site on the target sequence that is cleaved in a second chimeric IscB system mediated event.

[0548] In an embodiment, the recombination template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, e.g., one which results in the substitution of one amino acid for another in a protein product, e.g., transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation. In an embodiment, the recombination template nucleic acid can include sequence which results in an alteration in a non-coding sequence, e.g., an alteration in an exon

or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control element, e.g., a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

[0549] A recombination template nucleic acid having homology with a target position in a target gene may be used to alter the structure of a target sequence. The recombination template sequence may be used to alter an unwanted structure, e.g., an unwanted or mutant nucleotide. The recombination template nucleic acid may include sequence which, when integrated, results in: decreasing the activity of a positive control element; increasing the activity of a positive control element; decreasing the activity of a negative control element; increasing the activity of a negative control element; decreasing the expression of a gene; increasing the expression of a gene; increasing resistance to a disorder or disease; increasing resistance to viral entry; correcting a mutation or altering an unwanted amino acid residue conferring, increasing, abolishing or decreasing a biological property of a gene product, e.g., increasing the enzymatic activity of an enzyme, or increasing the ability of a gene product to interact with another molecule.

[0550] The recombination template nucleic acid may include sequence which results in: a change in sequence of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 1 1, 12 or more nucleotides of the target sequence. In an embodiment, the recombination template nucleic acid may be 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 1 10+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 1 80+/- 10, 190+/- 10, 200+/- 10, 210+/-10, of 220+/- 10 nucleotides in length. In an embodiment, the t recombination template nucleic acid may be 30+/-20, 40+/-20, 50+/-20, 60+/-20, 70+/- 20, 80+/-20, 90+/-20, 100+/- 20, 1 10+/-20, 120+/-20, 130+/-20, 140+/-20, I 50+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/- 20, 200+/-20, 210+/-20, of 220+/-20 nucleotides in length. In an embodiment, the recombination template nucleic acid is 10 to 1,000, 20 to 900, 30 to 800, 40 to 700, 50 to 600, 50 to 500, 50 to 400, 50 to300, 50 to 200, or 50 to 100 nucleotides in length.

[0551] A recombination template nucleic acid comprises the following components: [5' homology arm]-[replacement sequence]-[3' homology arm]. The homology arms provide for recombination into the chromosome, thus replacing the undesired element, e.g., a mutation or signature, with the replacement sequence. In an embodiment, the homology arms flank the most distal cleavage sites. In an embodiment, the 3' end of the 5' homology arm is the position next to the 5' end of the replacement sequence. In an embodiment, the 5' homology arm can

extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' from the 5' end of the replacement sequence. In an embodiment, the 5' end of the 3' homology arm is the position next to the 3' end of the replacement sequence. In an embodiment, the 3' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 3' from the 3' end of the replacement sequence.

[0552] In an embodiment, one or both homology arms may be shortened to avoid including certain sequence repeat elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In one embodiment, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

[0553] In an embodiment, a recombination template nucleic acids for correcting a mutation may designed for use as a single-stranded oligonucleotide. When using a single-stranded oligonucleotide, 5' and 3' homology arms may range up to about 200 base pairs (bp) in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length.

[0554] Mutating key residues in both DNA cleavage domains of the chimeric IscB system, results in the generation of a catalytically inactive chimeric IscB system. A catalytically inactive chimeric IscB system complexes with a guide RNA and localizes to the DNA sequence specified by that guide RNA's targeting domain, however, it does not cleave the target DNA. Fusion of the inactive chimeric IscB systems to an effector domain, e.g., a transcription repression domain, enables recruitment of the effector to any DNA site specified by the guide RNA. In an embodiment, chimeric IscB systems may be fused to a transcriptional repression domain and recruited to the promoter region of a gene. Especially for gene repression, it is contemplated herein that blocking the binding site of an endogenous transcription factor would aid in downregulating gene expression. In another embodiment, an inactive chimeric IscB systems can be fused to a chromatin modifying protein. Altering chromatin status can result in decreased expression of the target gene.

[0555] In an embodiment, a guide RNA molecule can be targeted to a known transcription response elements (e.g., promoters, enhancers, etc.), a known upstream activating sequences, and/or sequences of unknown or known function that are suspected of being able to control expression of the target DNA.

[0556] In some methods, a target polynucleotide can be inactivated to affect the modification of the expression in a cell. For example, upon the binding of a composition to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced.

Non-Homologous End-Joining

[0557] In an embodiment, nuclease-induced non-homologous end-joining (NHEJ) can be used to target gene-specific knockouts. Nuclease-induced NHEJ can also be used to remove (e.g., delete) sequence in a gene of interest. Generally, NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. Two-thirds of these mutations typically alter the reading frame and, therefore, produce a non-functional protein. Additionally, mutations that maintain the reading frame, but which insert or delete a significant amount of sequence, can destroy functionality of the protein. This is locus dependent as mutations in critical functional domains are likely less tolerable than mutations in non-critical regions of the protein. The indel mutations generated by NHEJ are unpredictable in nature; however, at a given break site certain indel sequences are favored and are over-represented in the population, likely due to small regions of microhomology. The lengths of deletions can vary widely; most commonly in the 1-50 bp range, but they can easily be greater than 50 bp, e.g., they can easily reach greater than about 100-200 bp. Insertions tend to be shorter and often include short duplications of the sequence immediately surrounding the break site. However, it is possible to obtain large insertions, and in these cases, the inserted sequence has often been traced to other regions of the genome or to plasmid DNA present in the cells.

[0558] Because NHEJ is a mutagenic process, it may also be used to delete small sequence motifs as long as the generation of a specific final sequence is not required. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger

DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences; however, the error-prone nature of NHEJ may still produce indel mutations at the site of repair.

[0559] Both double-strand cleaving chimeric IscB system, or an ortholog or homolog thereof, and single strand, or nickase, chimeric IscB system, or an ortholog or homolog thereof, molecules can be used in the methods and compositions described herein to generate NHEJ-mediated indels. NHEJ-mediated indels targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp).

[0560] In an embodiment, in which a guide RNA and chimeric IscB system, or an ortholog or homolog thereof, generate a double strand break for the purpose of inducing NHEJ-mediated indels, a guide RNA may be configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site may be between 0-500 bp away from the target position (e.g., less than 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position).

[0561] In an embodiment, in which two guide RNAs complexing with chimeric IscB system, or an ortholog or homolog thereof, e.g., IscB polypeptide nickases induce two single strand breaks for the purpose of inducing NHEJ-mediated indels, two guide RNAs may be configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position.

[0562] In some examples, the systems herein may introduce one or more indels via NHEJ pathway and insert sequence from a combination template via HDR.

EXEMPLARY APPLICATIONS

[0563] The invention provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or vector or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a modifying a target cell *in vivo*, *ex vivo* or *in vitro* and, may be conducted in a manner alters the cell such that once modified the progeny or cell line of the chimeric IscB

system modified cell retains the altered phenotype. The modified cells and progeny may be part of a multi-cellular organism such as a plant or animal with *ex vivo* or *in vivo* application of composition to desired cell types. The methods herein include a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or genome editing, or gene therapy.

[0564] In one embodiment, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In one embodiment, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. Methods for producing transgenic animals and plants are known in the art, and generally begin with a method of cell transfection, such as described herein.

Use of orthogonal catalytically inactive chimeric IscB systems

[0565] In one embodiment, the IscB polypeptide nickase, is used in combination with an orthogonal catalytically inactive IscB polypeptide, or chimeric IscB system nuclease to increase efficiency of said nickase (e.g., as described in Chen et al. 2017, Nature Communications 8:14958; doi:10.1038/ncomms14958). More particularly, the orthogonal catalytically inactive chimeric IscB system is characterized by a different TAM recognition site than the IscB nickase used in the AD-functionalized composition and the corresponding guide sequence is selected to bind to a target sequence proximal to that of the nickase of the functionalized chimeric IscB systems. The orthogonal catalytically inactive chimeric IscB systems as used in the context of the present invention does not form part of the functionalized composition but merely functions to increase the efficiency of said nickase and is used in combination with a standard ωRNA as described in the art for said chimeric IscB systems. In one embodiment, said orthogonal catalytically inactive chimeric IscB system is a dead chimeric IscB system, i.e., comprising one or more mutations which abolishes the nuclease activity of said chimeric IscB system. In one embodiment, the catalytically inactive orthogonal chimeric IscB system is provided with two or more ωRNAs or guide RNAs which are capable of hybridizing to target sequences which are proximal to the target sequence of the nickase. In one embodiment, at least two ωRNAs are used to target said catalytically inactive chimeric IscB systems, of which at least one ωRNA or guide RNAs is capable of hybridizing to a target sequence 5" of the target sequence of the nickase and at least one ωRNA is capable of hybridizing to a target sequence 3' of the target sequence of the nickase of the functionalized composition, whereby said one or more target sequences may be on the same or the opposite

DNA strand as the target sequence of the chimeric IscB system nickase. In one embodiment, the guide sequences for the one or more ωRNA s of the orthogonal catalytically inactive chimeric IscB systems are selected such that the target sequences are proximal to that of the ωRNA for the targeting of the functionalized composition, e.g., for the targeting of the nickase. In one embodiment, the one or more target sequences of the orthogonal catalytically inactive chimeric IscB systems are each separated from the target sequence of the nickase by more than 5 but less than 450 base pairs. Optimal distances between the target sequences of the guides for use with the orthogonal catalytically inactive chimeric IscB systems and the target sequence of the functionalized composition can be determined by the skilled person. In one embodiment, the catalytically inactive orthogonal chimeric IscB system has been modified to alter its TAM specificity as described elsewhere herein. In one embodiment, the chimeric IscB system nickase is a nickase which, by itself has limited activity in human cells, but which, in combination with an inactive orthogonal chimeric IscB system and one or more corresponding proximal guides ensures the required nickase activity.

Detection methods such as FISH

[0566] In one aspect, the invention provides an engineered, non-naturally occurring composition comprising a catalytically inactivate chimeric IscB system described herein and use this system in detection methods such as fluorescence in situ hybridization (FISH). A dead chimeric IscB system which lacks the ability to produce DNA double-strand breaks may be fused with a marker, such as fluorescent protein, such as the enhanced green fluorescent protein (eEGFP) and co-expressed with small guide RNAs to target pericentric, centric and teleomeric repeats *in vivo*. The dead IscB polypeptide or chimeric IscB systems can be used to visualize both repetitive sequences and individual genes in the human genome. Such new applications of labelled dead chimeric IscB system may be important in imaging cells and studying the functional nuclear architecture, especially in cases with a small nucleus volume or complex 3-D structures. (Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155(7):1479-91. doi: 10.1016/j.cell.2013.12.001.)

Patient-specific screening methods

[0567] A nucleic acid-targeting system that targets DNA, e.g., trinucleotide repeats can be used to screen patients or patent samples for the presence of such repeats. The repeats can be

the target of the RNA of the nucleic acid-targeting system, and if there is binding thereto by the nucleic acid-targeting system, that binding can be detected, to thereby indicate that such a repeat is present. Thus, a nucleic acid-targeting system can be used to screen patients or patient samples for the presence of the repeat. The patient can then be administered suitable compound(s) to address the condition; or can be administered a nucleic acid-targeting system to bind to and cause insertion, deletion or mutation and alleviate the condition.

Models of Genetic and Epigenetic Conditions

[0568] A method of the invention may be used to create a plant, an animal or cell that may be used to model and/or study genetic or epigenetic conditions of interest, such as a through a model of mutations of interest or a disease model. As used herein, "disease" refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more nucleic acid sequences associated with a disease, or a plant, animal or cell in which the expression of one or more nucleic acid sequences associated with a disease are altered. Such a nucleic acid sequence may encode a disease associated protein sequence or may be a disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism or cell can be a non-human subject, patient, organism or cell. Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Bacterial cell lines produced by the invention are also envisaged. Hence, cell lines are also envisaged.

[0569] In some methods, the disease model can be used to study the effects of mutations on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease.

[0570] In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated gene or polynucleotide can be modified such that the disease development and/or progression is inhibited or reduced. In particular, the

method comprises modifying a disease-associated gene or polynucleotide such that an altered protein is produced and, as a result, the animal or cell has an altered response. Accordingly, in some methods, a genetically modified animal may be compared with an animal predisposed to development of the disease such that the effect of the gene therapy event may be assessed.

[0571] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of a chimeric IscB system, and a conserved nucleotide sequence linked to a guide sequence; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[0572] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study the effects of a genome sequence modified by the complex of the invention on a cellular function of interest. For example, a cellular function model may be used to study the effect of a modified genome sequence on intracellular signaling or extracellular signaling. Alternatively, a cellular function model may be used to study the effects of a modified genome sequence on sensory perception. In some such models, one or more genome sequences associated with a signaling biochemical pathway in the model are modified.

[0573] Several disease models have been specifically investigated. These include de novo autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred but serve to show the broad applicability of the invention across genes and corresponding models. An altered expression of one or more genome sequences associated with a signaling biochemical pathway can be determined by assaying for a difference in the mRNA levels of the corresponding genes between the test model cell and a control cell, when they are contacted with a candidate agent. Alternatively, the differential expression of the sequences associated with a signaling biochemical pathway is determined by detecting a difference in the level of the encoded polypeptide or gene product.

[0574] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, nucleic acid contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes

or chemical solutions according to the procedures set forth in Sambrook et al. (1989) or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by amplification procedures or conventional hybridization assays (e.g., Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[0575] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of *E.coli* DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR. In particular, the isolated RNA can be subjected to a reverse transcription assay that is coupled with a quantitative polymerase chain reaction (RT-PCR) in order to quantify the expression level of a sequence associated with a signaling biochemical pathway.

[0576] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, one can conveniently determine the amount of the amplified products by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include SYBR green, SYBR blue, DAPI, propidium iodine, Hoeste, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and the like.

[0577] In another aspect, other fluorescent labels such as sequence specific probes can be employed in the amplification reaction to facilitate the detection and quantification of the amplified products. Probe-based quantitative amplification relies on the sequence-specific detection of a desired amplified product. It utilizes fluorescent, target-specific probes (e.g., TaqMan® probes) resulting in increased specificity and sensitivity. Methods for performing probe-based quantitative amplification are well established in the art and are taught in U.S. Patent No. 5,210,015.

[0578] In yet another aspect, conventional hybridization assays using hybridization probes that share sequence homology with sequences associated with a signaling biochemical pathway can be performed. Typically, probes are allowed to form stable complexes with the sequences associated with a signaling biochemical pathway contained within the biological sample derived from the test subject in a hybridization reaction. It will be appreciated by one of skill in the art that where antisense is used as the probe nucleic acid, the target polynucleotides provided in the sample are chosen to be complementary to sequences of the antisense nucleic acids. Conversely, where the nucleotide probe is a sense nucleic acid, the target polynucleotide is selected to be complementary to sequences of the sense nucleic acid.

[0579] Hybridization can be performed under conditions of various stringency. Suitable hybridization conditions for the practice of the present invention are such that the recognition interaction between the probe and sequences associated with a signaling biochemical pathway is both sufficiently specific and sufficiently stable. Conditions that increase the stringency of a hybridization reaction are widely known and published in the art. See, for example, (Sambrook, et al., (1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, second edition). The hybridization assay can be formed using probes immobilized on any solid support, including but are not limited to nitrocellulose, glass, silicon, and a variety of gene arrays. A preferred hybridization assay is conducted on high-density gene chips as described in U.S. Patent No. 5,445,934.

[0580] For a convenient detection of the probe-target complexes formed during the hybridization assay, the nucleotide probes are conjugated to a detectable label. Detectable labels suitable for use in the present invention include any composition detectable by photochemical, biochemical, spectroscopic, immunochemical, electrical, optical or chemical means. A wide variety of appropriate detectable labels are known in the art, which include fluorescent or chemiluminescent labels, radioactive isotope labels, enzymatic or other ligands. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as digoxigenin, ß-galactosidase, urease, alkaline phosphatase or peroxidase, avidin/biotin complex.

[0581] The detection methods used to detect or quantify the hybridization intensity will typically depend upon the label selected above. For example, radiolabels may be detected using photographic film or a phosphoimager. Fluorescent markers may be detected and quantified using a photodetector to detect emitted light. Enzymatic labels are typically detected by

providing the enzyme with a substrate and measuring the reaction product produced by the action of the enzyme on the substrate; and finally colorimetric labels are detected by simply visualizing the colored label.

[0582] An agent-induced change in expression of sequences associated with a signaling biochemical pathway can also be determined by examining the corresponding gene products. Determining the protein level typically involves a) contacting the protein contained in a biological sample with an agent that specifically bind to a protein associated with a signaling biochemical pathway; and (b) identifying any agent:protein complex so formed. In one aspect of this embodiment, the agent that specifically binds a protein associated with a signaling biochemical pathway is an antibody, preferably a monoclonal antibody.

[0583] The reaction is performed by contacting the agent with a sample of the proteins associated with a signaling biochemical pathway derived from the test samples under conditions that will allow a complex to form between the agent and the proteins associated with a signaling biochemical pathway. The formation of the complex can be detected directly or indirectly according to standard procedures in the art. In the direct detection method, the agents are supplied with a detectable label and unreacted agents may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. For such method, it is preferable to select labels that remain attached to the agents even during stringent washing conditions. It is preferable that the label does not interfere with the binding reaction. In the alternative, an indirect detection procedure may use an agent that contains a label introduced either chemically or enzymatically. A desirable label generally does not interfere with binding or the stability of the resulting agent:polypeptide complex. However, the label is typically designed to be accessible to an antibody for an effective binding and, hence generating a detectable signal.

[0584] A wide variety of labels suitable for detecting protein levels are known in the art. Non-limiting examples include radioisotopes, enzymes, colloidal metals, fluorescent compounds, bioluminescent compounds, and chemiluminescent compounds.

[0585] The amount of agent:polypeptide complexes formed during the binding reaction can be quantified by standard quantitative assays. As illustrated above, the formation of agent:polypeptide complex can be measured directly by the amount of label remained at the site of binding. In an alternative, the protein associated with a signaling biochemical pathway is tested for its ability to compete with a labeled analog for binding sites on the specific agent.

In this competitive assay, the amount of label captured is inversely proportional to the amount of protein sequences associated with a signaling biochemical pathway present in a test sample.

[0586] A number of techniques for protein analysis based on the general principles outlined above are available in the art. They include but are not limited to radioimmunoassay, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[0587] Antibodies that specifically recognize or bind to proteins associated with a signaling biochemical pathway are preferable for conducting the aforementioned protein analyses. Where desired, antibodies that recognize a specific type of post-translational modifications (e.g., signaling biochemical pathway inducible modifications) can be used. Post-translational modifications include but are not limited to glycosylation, lipidation, acetylation, and phosphorylation. These antibodies may be purchased from commercial vendors. For example, anti-phosphotyrosine antibodies that specifically recognize tyrosine-phosphorylated proteins are available from a number of vendors including Invitrogen and Perkin Elmer. Anti-phosphotyrosine antibodies are particularly useful in detecting proteins that are differentially phosphorylated on their tyrosine residues in response to an ER stress. Such proteins include but are not limited to eukaryotic translation initiation factor 2 alpha (eIF-2α). Alternatively, these antibodies can be generated using conventional polyclonal or monoclonal antibody technologies by immunizing a host animal or an antibody-producing cell with a target protein that exhibits the desired post-translational modification.

Genome Wide Knock-out Screening

[0588] The IscB polypeptide nuclease and systems described herein can be used to perform efficient and cost effective functional genomic screens. Such screens can utilize IscB polypeptide nuclease-based genome wide libraries. Such screens and libraries can provide for determining the function of genes, cellular pathways genes are involved in, and how any alteration in gene expression can result in a particular biological process. An advantage of the present invention is that the composition avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA. In preferred embodiments of the invention, the chimeric IscB system complexes are IscB polypeptide nuclease complexes.

[0589] In embodiments of the invention, a genome wide library may comprise a plurality of IscB polypeptide nuclease guide RNAs, as described herein, comprising guide sequences that are capable of targeting a plurality of target sequences in a plurality of genomic loci in a population of eukaryotic cells. The population of cells may be a population of embryonic stem (ES) cells. The target sequence in the genomic locus may be a non-coding sequence. The non-coding sequence may be an intron, regulatory sequence, splice site, 3' UTR, 5' UTR, or polyadenylation signal. Gene function of one or more gene products may be altered by said targeting. The targeting may result in a knockout of gene function. The targeting of a gene product may comprise more than one guide RNA. A gene product may be targeted by 2, 3, 4, 5, 6, 7, 8, 9, or 10 guide RNAs, preferably 3 to 4 per gene. Off-target modifications may be minimized by exploiting the staggered double strand breaks generated by IscB polypeptide nuclease complexes or by utilizing methods analogous to those used in composition (See, e.g., DNA targeting specificity of RNA-guided Cas nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013)), incorporated herein by reference. The targeting may be of about 100 or more sequences. The targeting may be of about 1000 or more sequences. The targeting may be of about 20,000 or more sequences. The targeting may be of the entire genome. The targeting may be of a panel of target sequences focused on a relevant or desirable pathway. The pathway may be an immune pathway. The pathway may be a cell division pathway.

[0590] One aspect of the invention comprehends a genome wide library that may comprise a plurality of ωRNAs or guide RNAs that may comprise guide sequences that are capable of targeting a plurality of target sequences in a plurality of genomic loci, wherein said targeting results in a knockout of gene function. This library may potentially comprise ωguide RNAs that target each and every gene in the genome of an organism.

[0591] In one embodiment of the invention, the organism or subject is a eukaryote (including mammal including human) or a non-human eukaryote or a non-human animal or a non-human mammal. In one embodiment, the organism or subject is a non-human animal, and may be an arthropod, for example, an insect, or may be a nematode. In some methods of the invention the organism or subject is a plant. In some methods of the invention, the organism or subject is a mammal or a non-human mammal. A non-human mammal may be for example a

rodent (preferably a mouse or a rat), an ungulate, or a primate. In some methods of the invention the organism or subject is algae, including microalgae, or is a fungus.

[0592] The knockout of gene function may comprise introducing into each cell in the population of cells a vector system of one or more vectors comprising an engineered, non-naturally occurring composition herein. The guide sequence may target a unique gene in each cell, wherein the chimeric IscB system is operably linked to a regulatory element, wherein when transcribed, the ωRNAs or guide RNA comprising the guide sequence directs sequence-specific binding of the IscB polypeptide nuclease to a target sequence in the genomic loci of the unique gene, inducing cleavage of the genomic loci by the chimeric IscB systems, and confirming different knockout mutations in a plurality of unique genes in each cell of the population of cells thereby generating a gene knockout cell library. The invention comprehends that the population of cells is a population of eukaryotic cells, and in a preferred embodiment, the population of cells is a population of embryonic stem (ES) cells.

[0593] The one or more vectors may be plasmid vectors. The vector may be a single vector comprising a chimeric IscB system, a ωRNA, and optionally, a selection marker into target cells. Not being bound by a theory, the ability to simultaneously deliver a chimeric IscB system and ωRNA through a single vector enables application to any cell type of interest, without the need to first generate cell lines that express the chimeric IscB systems. The regulatory element may be an inducible promoter. The inducible promoter may be a doxycycline inducible promoter. In some methods of the invention the expression of the guide sequence is under the control of the T7 promoter and is driven by the expression of T7 polymerase. The confirming of different knockout mutations may be by whole exome sequencing. The knockout mutation may be achieved in 100 or more unique genes. The knockout mutation may be achieved in 1000 or more unique genes. The knockout mutation may be achieved in 20,000 or more unique genes. The knockout mutation may be achieved in the entire genome. The knockout of gene function may be achieved in a plurality of unique genes which function in a particular physiological pathway or condition. The pathway or condition may be an immune pathway or condition. The pathway or condition may be a cell division pathway or condition.

[0594] Useful in the practice of the instant invention utilizing chimeric IscB system complexes are methods used in compositions and reference is made to: Genome-Scale CRISPR-Cas Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F.

Science Dec 12. (2013). [Epub ahead of print]; Published in final edited form as: Science. 2014 Jan 3; 343(6166): 84–87. Shalem et al. involves a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hitsNF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with IscB polypeptide nuclease.

[0595] Reference is also made to US patent publication number US20140357530; and PCT Patent Publication WO2014093701, hereby incorporated herein by reference. Reference is also made to NIH Press Release of Oct. 22, 2015 entitled, "Researchers identify potential alternative to CRISPR-Cas genome editing tools: New Cas enzymes shed light on evolution of CRISPR-Cas systems, which is incorporated by reference.

Functional Alteration and Screening

[0596] In another aspect, the present invention provides for a method of functional evaluation and screening of genes. The use of the compositions to precisely deliver functional domains, to activate or repress genes or to alter epigenetic state by precisely altering the methylation site on a specific locus of interest, can be with one or more ωRNAs or guide RNAs applied to a single cell or population of cells or with a library applied to genome in a pool of cells *ex vivo* or *in vivo* comprising the administration or expression of a library comprising a plurality of ωRNA s (comprising guide molecules) and wherein the screening further comprises use of a chimeric IscB systems, wherein the complex comprising the chimeric IscB system is modified to comprise a heterologous functional domain. In an aspect the invention provides a method for screening a genome comprising the administration to a host or expression in a host *in vivo* of a library. In an aspect the invention provides a method as herein discussed further comprising an activator administered to the host or expressed in the host. In an aspect the invention provides a method as herein discussed wherein the activator is attached to a chimeric

IscB system. In an aspect the invention provides a method as herein discussed wherein the activator is attached to the N terminus or the C terminus of the chimeric IscB systems. In an aspect the invention provides a method as herein discussed wherein the activator is attached to a ωRNA loop. In an aspect the invention provides a method as herein discussed further comprising a repressor administered to the host or expressed in the host. In an aspect the invention provides a method as herein discussed, wherein the screening comprises affecting and detecting gene activation, gene inhibition, or cleavage in the locus.

[0597] It is also preferred to target endogenous (regulatory) control elements (such as enhancers and silencers) e.g. in addition to a promoter or promoter-proximal elements. Thus, the invention can also be used to target endogenous control elements (including enhancers and silencers) in addition to targeting of the promoter. These control elements can be located upstream and downstream of the transcriptional start site (TSS), starting from 200bp from the TSS to 100kb away. Targeting of known control elements can be used to activate or repress the gene of interest. In some cases, a single control element can influence the transcription of multiple target genes. Targeting of a single control element could therefore be used to control the transcription of multiple genes simultaneously.

[0598] Targeting of putative control elements on the other hand (e.g., by tiling the region of the putative control element as well as 200bp up to 100kB around the element) can be used as a means to verify such elements (by measuring the transcription of the gene of interest) or to detect novel control elements (e.g., by tiling 100kb upstream and downstream of the TSS of the gene of interest). In addition, targeting of putative control elements can be useful in the context of understanding genetic causes of disease. Many mutations and common SNP variants associated with disease phenotypes are located outside coding regions. Targeting of such regions with either the activation or repression systems described herein can be followed by readout of transcription of either a) a set of putative targets (e.g., a set of genes located in closest proximity to the control element) or b) whole-transcriptome readout by e.g., RNAseq or microarray. This would allow for the identification of likely candidate genes involved in the disease phenotype. Such candidate genes could be useful as novel drug targets.

[0599] Histone acetyltransferase (HAT) inhibitors are mentioned herein. However, an alternative In one embodiment is for the one or more functional domains to comprise an acetyltransferase, preferably a histone acetyltransferase. These are useful in the field of epigenomics, for example in methods of interrogating the epigenome. Methods of interrogating

the epigenome may include, for example, targeting epigenomic sequences. Targeting epigenomic sequences may include the guide being directed to an epigenomic target sequence. Epigenomic target sequence may include, in one embodiment, include a promoter, silencer or an enhancer sequence.

Saturating Mutagenesis

[0600] The compositions herein can be used to perform saturating or deep scanning mutagenesis of genomic loci in conjunction with a cellular phenotype—for instance, for determining critical minimal features and discrete vulnerabilities of functional elements required for gene expression, drug resistance, and reversal of disease. By saturating or deep scanning mutagenesis is meant that every or essentially every DNA base is cut within the genomic loci. A library of Cas1 effector protein guide RNAs may be introduced into a population of cells. The library may be introduced, such that each cell receives a single ωRNA. In the case where the library is introduced by transduction of a viral vector, as described herein, a low multiplicity of infection (MOI) is used. The library may include ωRNAs targeting every sequence upstream of a TAM sequence in a genomic locus. The library may include at least 100 non-overlapping genomic sequences upstream of a TAM sequence for every 1000 base pairs within the genomic locus. The library may include ωRNAs targeting sequences upstream of at least one different TAM sequence. The composition may include more than one chimeric IscB system. Any chimeric IscB systems as described herein, including orthologues or engineered chimeric IscB systems that recognize different TAM sequences may be used. The frequency of off target sites for a ωRNA may be less than 500. Off target scores may be generated to select ωRNAs with the lowest off target sites. Any phenotype determined to be associated with cutting at a ωRNA target site may be confirmed by using ωRNAs targeting the same site in a single experiment. Validation of a target site may also be performed by using a modified chimeric IscB system, as described herein, and two ωRNAs targeting the genomic site of interest. Not being bound by a theory, a target site is a true hit if the change in phenotype is observed in validation experiments.

[0601] The genomic loci may include at least one continuous genomic region. The at least one continuous genomic region may comprise up to the entire genome. The at least one continuous genomic region may comprise a functional element of the genome. The functional element may be within a non-coding region, coding gene, intronic region, promoter, or enhancer. The at least one continuous genomic region may comprise at least 1 kb, preferably

at least 50 kb of genomic DNA. The at least one continuous genomic region may comprise a transcription factor binding site. The at least one continuous genomic region may comprise a region of DNase I hypersensitivity. The at least one continuous genomic region may comprise a transcription enhancer or repressor element. The at least one continuous genomic region may comprise a site enriched for an epigenetic signature. The at least one continuous genomic DNA region may comprise an epigenetic insulator. The at least one continuous genomic region may comprise two or more continuous genomic regions that physically interact. Genomic regions that interact may be determined by '4C technology'. 4C technology allows the screening of the entire genome in an unbiased manner for DNA segments that physically interact with a DNA fragment of choice, as is described in Zhao et al. ((2006) Nat Genet 38, 1341-7) and in U.S. patent 8,642,295, both incorporated herein by reference in its entirety. The epigenetic signature may be histone acetylation, histone methylation, histone ubiquitination, histone phosphorylation, DNA methylation, or a lack thereof.

[0602] The compositions for saturating or deep scanning mutagenesis can be used in a population of cells. The compositions can be used in eukaryotic cells, including but not limited to mammalian and plant cells. The population of cells may be prokaryotic cells. The population of eukaryotic cells may be a population of embryonic stem (ES) cells, neuronal cells, epithelial cells, immune cells, endocrine cells, muscle cells, erythrocytes, lymphocytes, plant cells, or yeast cells.

[0603] In one aspect, the present invention provides for a method of screening for functional elements associated with a change in a phenotype. The library may be introduced into a population of cells that are adapted to contain a chimeric IscB system. The cells may be sorted into at least two groups based on the phenotype. The phenotype may be expression of a gene, cell growth, or cell viability. The relative representation of the ωRNAs or guide RNAs present in each group are determined, whereby genomic sites associated with the change in phenotype are determined by the representation of ωRNAs or guide RNAs present in each group. The change in phenotype may be a change in expression of a gene of interest. The gene of interest may be upregulated, downregulated, or knocked out. The cells may be sorted into a high expression group and a low expression group. The population of cells may include a reporter construct that is used to determine the phenotype. The reporter construct may include a detectable marker. Cells may be sorted by use of the detectable marker.

[0604] In another aspect, the present invention provides for a method of screening for genomic sites associated with resistance to a chemical compound. The chemical compound may be a drug or pesticide. The library may be introduced into a population of cells that are adapted to contain a chimeric IscB system, wherein each cell of the population contains no more than one ωRNA or guide RNA; the population of cells are treated with the chemical compound; and the representation of ωRNA or guide RNAs are determined after treatment with the chemical compound at a later time point as compared to an early time point, whereby genomic sites associated with resistance to the chemical compound are determined by enrichment of ωRNAs. Representation of ωRNAs may be determined by deep sequencing methods.

[0605] Useful in the practice of the instant invention utilizing compositions are methods used in compositions and reference is made to the article entitled BCL11A enhancer dissection by Cas-mediated in situ saturating mutagenesis. Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., Yuan, G., Zhang, F., Orkin, S.H., & Bauer, D.E. DOI:10.1038/nature15521, published online September 16, 2015, the article is herein incorporated by reference and discussed briefly below:

[0606] Canver et al. involves novel pooled guide RNA libraries to perform in situ saturating mutagenesis of the human and mouse BCL11A erythroid enhancers previously identified as an enhancer associated with fetal hemoglobin (HbF) level and whose mouse ortholog is necessary for erythroid BCL11A expression. This approach revealed critical minimal features and discrete vulnerabilities of these enhancers. Through editing of primary human progenitors and mouse transgenesis, the authors validated the BCL11A erythroid enhancer as a target for HbF reinduction. The authors generated a detailed enhancer map that informs therapeutic genome editing.

Modification of a Cell or Organism

[0607] The present disclosure further provides cells comprising one or more components of the systems herein, e.g., the chimeric IscB system(s) and/or ωRNA(s). Also provided include cells modified by the systems and methods herein, and cell cultures, tissues, organs, organism comprising such cells or progeny thereof. The invention In one embodiment comprehends a method of modifying a cell or organism. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The mammalian cell many be a non-human primate,

bovine, porcine, rodent or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry, fish or shrimp. The cell may also be a plant cell. The plant cell may be of a crop plant such as cassava, corn, sorghum, wheat, or rice. The plant cell may also be of an algae, tree or vegetable. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol or other desired cellular output. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell include an alteration that changes the biologic product produced.

THERAPEUTIC USES AND METHODS OF TREATMENT

[0608] Also provided herein are methods of diagnosing, prognosing, treating, and/or preventing a disease, state, or condition in or of a subject. Generally, the methods of diagnosing, prognosing, treating, and/or preventing a disease, state, or condition in or of a subject can include modifying a polynucleotide in a subject or cell thereof using a composition, system, or component thereof described herein and/or include detecting a diseased or healthy polynucleotide in a subject or cell thereof using a composition, system, or component thereof described herein. In one embodiment, the method of treatment or prevention can include using a composition, system, or component thereof to modify a polynucleotide of an infectious organism (e.g., bacterial or virus) within a subject or cell thereof. In one embodiment, the method of treatment or prevention can include using a composition, system, or component thereof to modify a polynucleotide of an infectious organism or symbiotic organism within a subject. The composition, system, and components thereof can be used to develop models of diseases, states, or conditions. The composition, system, and components thereof can be used to detect a disease state or correction thereof, such as by a method of treatment or prevention described herein. The composition, system, and components thereof can be used to screen and select cells that can be used, for example, as treatments or preventions described herein. The composition, system, and components thereof can be used to develop biologically active agents that can be used to modify one or more biologic functions or activities in a subject or a cell thereof.

[0609] In general, the method can include delivering a composition, system, and/or component thereof to a subject or cell thereof, or to an infectious or symbiotic organism by a suitable delivery technique and/or composition. Once administered the components can operate as described elsewhere herein to elicit a nucleic acid modification event. In some aspects, the

nucleic acid modification event can occur at the genomic, epigenomic, and/or transcriptomic level. DNA and/or RNA cleavage, gene activation, and/or gene deactivation can occur. Additional features, uses, and advantages are described in greater detail below. On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. In addition to treating and/or preventing a disease in a subject, the compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling *in vivo* (e.g., gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[0610] The composition, system, and components thereof described elsewhere herein can be used to treat and/or prevent a disease, such as a genetic and/or epigenetic disease, in a subject. The composition, system, and components thereof described elsewhere herein can be used to treat and/or prevent genetic infectious diseases in a subject, such as bacterial infections, viral infections, fungal infections, parasite infections, and combinations thereof. The composition, system, and components thereof described elsewhere herein can be used to modify the composition or profile of a microbiome in a subject, which can in turn modify the health status of the subject. The composition, system, described herein can be used to modify cells *ex vivo*, which can then be administered to the subject whereby the modified cells can treat or prevent a disease or symptom thereof. This is also referred to in some contexts as adoptive therapy. The composition, system, described herein can be used to treat mitochondrial diseases, where the mitochondrial disease etiology involves a mutation in the mitochondrial DNA.

[0611] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the polynucleotide encoding one or more components of the composition, system, or complex or any of polynucleotides or vectors described herein and administering them to the subject. A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. The repair template may be a recombination template herein. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or

repression of multiple target gene loci by transforming the subject with the polynucleotides or vectors described herein, wherein said polynucleotide or vector encodes or comprises one or more components of composition, system, complex or component thereof comprising multiple chimeric IscB systems. Where any treatment is occurring *ex vivo*, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[0612] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the chimeric IscB system(s), advantageously encoding and expressing *in vivo* the remaining portions of the composition, system, (e.g., RNA, guides). A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or repression by transforming the subject with the chimeric IscB system(s) advantageously encoding and expressing *in vivo* the remaining portions of the composition, system, (e.g., RNA, guides); advantageously. In one embodiment, the chimeric IscB system is a catalytically inactive chimeric IscB system and includes one or more associated functional domains. Where any treatment is occurring *ex vivo*, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[0613] One or more components of the composition and system described herein can be included in a composition, such as a pharmaceutical composition, and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g., lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g., for lentiviral ωRNA selection) and concentration of ωRNA (e.g. dependent on whether multiple ωRNAs are used) may be advantageous for eliciting an improved effect.

[0614] Thus, also described herein are methods of inducing one or more polynucleotide modifications in a eukaryotic or prokaryotic cell or component thereof (e.g., a mitochondria) of a subject, infectious organism, and/or organism of the microbiome of the subject. The modification can include the introduction, deletion, or substitution of one or more nucleotides at a target sequence of a polynucleotide of one or more cell(s). The modification can occur *in vitro*, *ex vivo*, *in situ*, or *in vivo*.

[0615] In one embodiment, the method of treating or inhibiting a condition or a disease caused by one or more mutations in a genomic locus in a eukaryotic organism or a non-human organism can include manipulation of a target sequence within a coding, non-coding or regulatory element of said genomic locus in a target sequence in a subject or a non-human subject in need thereof comprising modifying the subject or a non-human subject by manipulation of the target sequence and wherein the condition or disease is susceptible to treatment or inhibition by manipulation of the target sequence including providing treatment comprising delivering a composition comprising the particle delivery system or the delivery system or the virus particle of any one of the above embodiment or the cell of any one of the above embodiment.

[0616] Also provided herein is the use of the particle delivery system or the delivery system or the virus particle of any one of the above embodiments or the cell of any one of the above embodiment in *ex vivo* or *in vivo* gene or genome editing; or for use in *in vitro*, *ex vivo* or *in vivo* gene therapy. Also provided herein are particle delivery systems, non-viral delivery systems, and/or the virus particle of any one of the above embodiments or the cell of any one of the above embodiments used in the manufacture of a medicament for *in vitro*, *ex vivo* or *in vivo* gene or genome editing or for use in *in vitro*, *ex vivo* or *in vivo* gene therapy or for use in a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus associated with a disease or in a method of treating or inhibiting a condition or disease caused by one or more mutations in a genomic locus in a eukaryotic organism or a non- human organism.

[0617] In one embodiment, polynucleotide modification can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said polynucleotide of said cell(s). The modification can include the introduction, deletion, or substitution of at least 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence. The modification can include the introduction, deletion, or substitution of at least 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35,

40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, 5600, 5700, 5800, 5900, 6000, 6100, 6200, 6300, 6400, 6500, 6600, 6700, 6800, 6900, 7000, 7100, 7200, 7300, 7400, 7500, 7600, 7700, 7800, 7900, 8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500, 9600, 9700, 9800, or 9900 to 10000 nucleotides at each target sequence of said cell(s).

[0618] In one embodiment, the modifications can include the introduction, deletion, or substitution of nucleotides at each target sequence of said cell(s) via nucleic acid components (e.g., guide(s) RNA(s) or ωRNA(s)), such as those mediated by a composition, system, or a component thereof described elsewhere herein. In one embodiment, the modifications can include the introduction, deletion, or substitution of nucleotides at a target or random sequence of said cell(s) via a composition, system, or technique.

[0619] In one embodiment, the composition, system, or component thereof can promote Non-Homologous End-Joining (NHEJ). In one embodiment, modification of a polynucleotide by a composition, system, or a component thereof, such as a diseased polynucleotide, can include NHEJ. In one embodiment, promotion of this repair pathway by the composition, system, or a component thereof can be used to target gene or polynucleotide specific knock-outs and/or knock-ins. In one embodiment, promotion of this repair pathway by the composition, system, or a component thereof can be used to generate NHEJ-mediated indels. Nuclease-induced NHEJ can also be used to remove (e.g., delete) sequence in a gene of interest. Generally, NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. The indel can range in size from 1-50 or more base pairs. In one embodiment

thee indel can be 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, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 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, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, or 500 base pairs or more. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences.

[0620] In one embodiment, composition, system, mediated NHEJ can be used in the method to delete small sequence motifs. In one embodiment, composition, system, mediated NHEJ can be used in the method to generate NHEJ-mediate indels that can be targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp). In an embodiment, in which a ωRNA or guide RNA and chimeric IscB system generate a double strand break for the purpose of inducing NHEJ-mediated indels, a ωRNA or guide RNA may be configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site may be between 0-500 bp away from the target position (e.g., less than 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position). In an embodiment, in which two ωRNA or guide RNAs complexing with one or more nickases induce two single strand breaks for the purpose of inducing NHEJ-mediated indels, two guide RNAs may be configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position.

[0621] For minimization of toxicity and off-target effect, it may be important to control the concentration of chimeric IscB system mRNA and guide RNA delivered. Optimal concentrations of chimeric IscB system mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, nickase mRNA (for example S. pyogenes Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in International Patent Publication No. WO 2014/093622 (PCT/US2013/074667); or, via mutation. Others are as described elsewhere herein.

[0622] Typically, in the context of an endogenous chimeric IscB system, formation of a chimeric IscB system or complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more chimeric IscB systems) results in cleavage, nicking, and/or another modification of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence.

[0623] In one embodiment, a method of modifying a target polynucleotide in a cell to treat or prevent a disease can include allowing a composition, system, or component thereof to bind to the target polynucleotide, e.g., to effect cleavage, nicking, or other modification as the composition, system, is capable of said target polynucleotide, thereby modifying the target polynucleotide, wherein the composition, system, or component thereof, complex with a guide sequence, and hybridize said guide sequence to a target sequence within the target polynucleotide, wherein said guide sequence is optionally linked to a ωRNA scaffold sequence. In some of these embodiments, the composition, system, or component thereof can be or include a chimeric IscB system complexed with a guide sequence. In one embodiment, modification can include cleaving or nicking one or two strands at the location of the target sequence by one or more components of the composition, system, or component thereof.

[0624] The cleavage, nicking, or other modification capable of being performed by the composition, system, can modify transcription of a target polynucleotide. In one embodiment, modification of transcription can include decreasing transcription of a target polynucleotide. In one embodiment, modification can include increasing transcription of a target polynucleotide. In one embodiment, the method includes repairing said cleaved target polynucleotide by homologous recombination with a recombination template polynucleotide, wherein said repair results in a modification such as, but not limited to, an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In one embodiment, said modification results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In one embodiment, the modification imparted by the composition, system, or component thereof provides a transcript and/or protein that can correct a disease or a symptom thereof, including but not limited to, any of those described in greater detail elsewhere herein.

[0625] In one embodiment, the method of treating or preventing a disease can include delivering one or more vectors or vector systems to a cell, such as a eukaryotic or prokaryotic cell, wherein one or more vectors or vector systems include the composition, system, or component thereof. In one embodiment, the vector(s) or vector system(s) can be a viral vector or vector system, such as an AAV or lentiviral vector system, which are described in greater detail elsewhere herein. In one embodiment, the method of treating or preventing a disease can include delivering one or more viral particles, such as an AAV or lentiviral particle, containing the composition, system, or component thereof. In one embodiment, the viral particle has a tissue specific tropism. In one embodiment, the viral particle has a liver, muscle, eye, heart,

pancreas, kidney, neuron, epithelial cell, endothelial cell, astrocyte, glial cell, immune cell, or red blood cell specific tropism.

[0626] It will be understood that the composition and system, according to the invention as described herein, such as the composition and system, for use in the methods according to the invention as described herein, may be suitably used for any type of application known for composition, system, preferably in eukaryotes. In certain aspects, the application is therapeutic, preferably therapeutic in a eukaryote organism, such as including but not limited to animals (including human), plants, algae, fungi (including yeasts), etc. Alternatively, or in addition, in certain aspects, the application may involve accomplishing or inducing one or more particular traits or characteristics, such as genotypic and/or phenotypic traits or characteristics, as also described elsewhere herein.

Treating Diseases of the Circulatory System

[0627] In one embodiment, the composition, system, and/or component thereof described herein can be used to treat and/or prevent a circulatory system disease. Exemplary disease is provided, for example, in Table 6. In one embodiment the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) can be used to deliver the composition, system, and/or component thereof described herein to the blood. In one embodiment, the circulatory system disease can be treated by using a lentivirus to deliver the composition, system, described herein to modify hematopoietic stem cells (HSCs) *in vivo* or *ex vivo* (see e.g. Drakopoulou, "Review Article, The Ongoing Challenge of Hematopoietic Stem Cell-Based Gene Therapy for β-Thalassemia," Stem Cells International, Volume 2011, Article ID 987980, 10 pages, doi:10.4061/2011/987980, which can be adapted for use with the composition, system, herein in view of the description herein). In one embodiment, the circulatory system disorder can be treated by correcting HSCs as to the disease using a composition, system, herein or a component thereof, wherein the composition, system, optionally includes a suitable HDR repair template (see e.g. Cavazzana, "Outcomes of Gene Therapy for β-Thalassemia Major via Transplantation of Autologous Hematopoietic Stem Cells Transduced *Ex Vivo* with a Lentiviral βA-T87Q-Globin Vector."; Cavazzana-Calvo, "Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia", Nature 467, 318–322 (16 September 2010) doi:10.1038/nature09328; Nienhuis, "Development of Gene Therapy for Thalassemia, Cold Spring Harbor Perspectives in Medicine, doi: 10.1101/cshperspect.a011833 (2012), LentiGlobin BB305, a lentiviral vector containing an engineered β-globin gene (βA-T87Q);

and Xie et al., "Seamless gene correction of β-thalassaemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback" Genome Research gr.173427.114 (2014) www.genome.org/cgi/doi/10.1101/gr.173427.114 (Cold Spring Harbor Laboratory Press; [1599] Watts, "Hematopoietic Stem Cell Expansion and Gene Therapy" Cytotherapy 13(10):1164–1171. doi:10.3109/14653249.2011.620748 (2011), which can be adapted for use with the composition, system, herein in view of the description herein). In one embodiment, iPSCs can be modified using a composition, system, described herein to correct a disease polynucleotide associated with a circulatory disease. In this regard, the teachings of Xu et al. (Sci Rep. 2015 Jul 9;5:12065. doi: 10.1038/srep12065) and Song et al. (Stem Cells Dev. 2015 May 1;24(9):1053-65. doi: 10.1089/scd.2014.0347. Epub 2015 Feb 5) with respect to modifying iPSCs can be adapted for use in view of the description herein with the composition, system, described herein.

[0628] The term "Hematopoietic Stem Cell" or "HSC" refers broadly those cells considered to be an HSC, e.g., blood cells that give rise to all the other blood cells and are derived from mesoderm; located in the red bone marrow, which is contained in the core of most bones. HSCs of the invention include cells having a phenotype of hematopoietic stem cells, identified by small size, lack of lineage (lin) markers, and markers that belong to the cluster of differentiation series, like: CD34, CD38, CD90, CD133, CD105, CD45, and also c-kit, - the receptor for stem cell factor. Hematopoietic stem cells are negative for the markers that are used for detection of lineage commitment, and are, thus, called Lin-; and, during their purification by FACS, a number of up to 14 different mature blood-lineage markers, e.g., CD13 & CD33 for myeloid, CD71 for erythroid, CD19 for B cells, CD61 for megakaryocytic, etc. for humans; and, B220 (murine CD45) for B cells, Mac-1 (CD11b/CD18) for monocytes, Gr-1 for Granulocytes, Ter119 for erythroid cells, Il7Ra, CD3, CD4, CD5, CD8 for T cells, etc. Mouse HSC markers: CD34lo/-, SCA-1+, Thy1.1+/lo, CD38+, C-kit+, lin-, and Human HSC markers: CD34+, CD59+, Thy1/CD90+, CD38lo/-, C-kit/CD117+, and lin-. HSCs are identified by markers. Hence in embodiments discussed herein, the HSCs can be CD34+ cells. HSCs can also be hematopoietic stem cells that are CD34-/CD38-. Stem cells that may lack c-kit on the cell surface that are considered in the art as HSCs are within the ambit of the invention, as well as CD133+ cells likewise considered HSCs in the art.

[0629] In one embodiment, the treatment or prevention for treating a circulatory system or blood disease can include modifying a human cord blood cell with any modification described

herein. In one embodiment, the treatment or prevention for treating a circulatory system or blood disease can include modifying a granulocyte colony-stimulating factor-mobilized peripheral blood cell (mPB) with any modification described herein. In one embodiment, the human cord blood cell or mPB can be CD34+. In one embodiment, the cord blood cell(s) or mPB cell(s) modified can be autologous. In one embodiment, the cord blood cell(s) or mPB cell(s) can be allogenic. In addition to the modification of the disease gene(s), allogenic cells can be further modified using the composition, system, described herein to reduce the immunogenicity of the cells when delivered to the recipient. Such techniques are described elsewhere herein and e.g., Cartier, "MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy," Brain Pathology 20 (2010) 857–862, which can be adapted for use with the composition, system, herein. The modified cord blood cell(s) or mPB cell(s) can be optionally expanded *in vitro*. The modified cord blood cell(s) or mPB cell(s) can be derived to a subject in need thereof using any suitable delivery technique.

[0630] The compositions may be engineered to target genetic locus or loci in HSCs. In one embodiment, the chimeric IscB system(s) can be codon-optimized for a eukaryotic cell and especially a mammalian cell, e.g., a human cell, for instance, HSC, or iPSC and ωRNA targeting a locus or loci in HSC, such as circulatory disease, can be prepared. These may be delivered via particles. The particles may be formed by the chimeric IscB system and the ωRNA being admixed. The ωRNA and chimeric IscB system mixture can be, for example, admixed with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol, whereby particles containing the ωRNA and chimeric IscB systems may be formed. The invention comprehends so making particles and particles from such a method as well as uses thereof. Particles suitable delivery of the composition in the context of blood or circulatory system or HSC delivery to the blood or circulatory system are described in greater detail elsewhere herein.

[0631] In one embodiment, after *ex vivo* modification the HSCs or iPCS can be expanded prior to administration to the subject. Expansion of HSCs can be via any suitable method such as that described by, Lee, "Improved *ex vivo* expansion of adult hematopoietic stem cells by overcoming CUL4-mediated degradation of HOXB4." Blood. 2013 May 16;121(20):4082-9. doi: 10.1182/blood-2012-09-455204. Epub 2013 Mar 21.

[0632] In one embodiment, the HSCs or iPSCs modified can be autologous. In one embodiment, the HSCs or iPSCs can be allogenic. In addition to the modification of the disease gene(s), allogenic cells can be further modified using the composition, system, described herein to reduce the immunogenicity of the cells when delivered to the recipient. Such techniques are described elsewhere herein and e.g., Cartier, "MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy," Brain Pathology 20 (2010) 857–862, which can be adapted for use with the composition, system, herein.

Treating Neurological Diseases

[0633] In one embodiment, the compositions, systems, described herein can be used to treat diseases of the brain and CNS. Delivery options for the brain include encapsulation of chimeric IscB system and guide RNA in the form of either DNA or RNA into liposomes and conjugating to molecular Trojan horses for trans-blood brain barrier (BBB) delivery. Molecular Trojan horses have been shown to be effective for delivery of B-gal expression vectors into the brain of non-human primates. The same approach can be used to delivery vectors containing chimeric IscB system and guide RNA. For instance, Xia CF and Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology." Mol Pharm. 2009 May-Jun;6(3):747-51. doi: 10.1021/mp800194) describes how delivery of short interfering RNA (siRNA) to cells in culture, and *in vivo*, is possible with combined use of a receptor-specific monoclonal antibody (mAb) and avidin-biotin technology. The authors also report that because the bond between the targeting mAb and the siRNA is stable with avidin-biotin technology, and RNAi effects at distant sites such as brain are observed *in vivo* following an intravenous administration of the targeted siRNA, the teachings of which can be adapted for use with the compositions, systems, herein. In other embodiments, an artificial virus can be generated for CNS and/or brain delivery. See e.g., Zhang et al. (Mol Ther. 2003 Jan;7(1):11-8.)), the teachings of which can be adapted for use with the compositions, systems, herein.

Treating Hearing Diseases

[0634] In one embodiment the composition and system described herein can be used to treat a hearing disease or hearing loss in one or both ears. Deafness is often caused by lost or damaged hair cells that cannot relay signals to auditory neurons. In such cases, cochlear implants may be used to respond to sound and transmit electrical signals to the nerve cells. But

these neurons often degenerate and retract from the cochlea as fewer growth factors are released by impaired hair cells.

[0635] In one embodiment, the composition, system, or modified cells can be delivered to one or both ears for treating or preventing hearing disease or loss by any suitable method or technique. Suitable methods and techniques include, but are not limited to, those set forth in US Patent Publication No. 20120328580 describes injection of a pharmaceutical composition into the ear (e.g., auricular administration), such as into the luminae of the cochlea (e.g., the Scala media, Sc vestibulae, and Sc tympani), e.g., using a syringe, e.g., a single-dose syringe. For example, one or more of the compounds described herein can be administered by intratympanic injection (e.g., into the middle ear), and/or injections into the outer, middle, and/or inner ear; administration in situ, via a catheter or pump (see e.g. McKenna et al., (U.S. Patent Publication No. 2006/0030837) and Jacobsen et al., (U.S. Pat. No. 7,206,639); administration in combination with a mechanical device such as a cochlear implant or a hearing aid, which is worn in the outer ear (see e.g. U.S. Patent Publication No. 2007/0093878, which provides an exemplary cochlear implant suitable for delivery of the compositions, systems, described herein to the ear). Such methods are routinely used in the art, for example, for the administration of steroids and antibiotics into human ears. Injection can be, for example, through the round window of the ear or through the cochlear capsule. Other inner ear administration methods are known in the art (see, e.g., Salt and Plontke, Drug Discovery Today, 10:1299-1306, 2005). In one embodiment, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient during a surgical procedure. In one embodiment, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient without the need for a surgical procedure.

[0636] In general, the cell therapy methods described in US Patent Publication No. 20120328580 can be used to promote complete or partial differentiation of a cell to or towards a mature cell type of the inner ear (e.g., a hair cell) *in vitro*. Cells resulting from such methods can then be transplanted or implanted into a patient in need of such treatment. The cell culture methods required to practice these methods, including methods for identifying and selecting suitable cell types, methods for promoting complete or partial differentiation of selected cells, methods for identifying complete or partially differentiated cell types, and methods for implanting complete or partially differentiated cells are described below.

[0637] Cells suitable for use in the present invention include, but are not limited to, cells that are capable of differentiating completely or partially into a mature cell of the inner ear, e.g., a hair cell (e.g., an inner and/or outer hair cell), when contacted, e.g., *in vitro*, with one or more of the compounds described herein. Exemplary cells that are capable of differentiating into a hair cell include, but are not limited to stem cells (e.g., inner ear stem cells, adult stem cells, bone marrow derived stem cells, embryonic stem cells, mesenchymal stem cells, skin stem cells, iPS cells, and fat derived stem cells), progenitor cells (e.g., inner ear progenitor cells), support cells (e.g., Deiters' cells, pillar cells, inner phalangeal cells, tectal cells and Hensen's cells), and/or germ cells. The use of stem cells for the replacement of inner ear sensory cells is described in Li et al., (U.S. Patent Publication No. 2005/0287127) and Li et al., (U.S. Patent Application No. 11/953,797). The use of bone marrow derived stem cells for the replacement of inner ear sensory cells is described in Edge et al., PCT/US2007/084654. iPS cells are described, e.g., at Takahashi et al., Cell, Volume 131, Issue 5, Pages 861-872 (2007); Takahashi and Yamanaka, Cell 126, 663-76 (2006); Okita et al., Nature 448, 260-262 (2007); Yu, J. et al., Science 318(5858):1917-1920 (2007); Nakagawa et al., Nat. Biotechnol. 26:101-106 (2008); and Zaehres and Scholer, Cell 131(5):834-835 (2007). Such suitable cells can be identified by analyzing (e.g., qualitatively or quantitatively) the presence of one or more tissue specific genes. For example, gene expression can be detected by detecting the protein product of one or more tissue-specific genes. Protein detection techniques involve staining proteins (e.g., using cell extracts or whole cells) using antibodies against the appropriate antigen. In this case, the appropriate antigen is the protein product of the tissue-specific gene expression. Although, in principle, a first antibody (i.e., the antibody that binds the antigen) can be labeled, it is more common (and improves the visualization) to use a second antibody directed against the first (e.g., an anti-IgG). This second antibody is conjugated either with fluorochromes, or appropriate enzymes for colorimetric reactions, or gold beads (for electron microscopy), or with the biotin-avidin system, so that the location of the primary antibody, and thus the antigen, can be recognized.

[0638] The composition and system may be delivered to the ear by direct application of pharmaceutical composition to the outer ear, with compositions modified from US Patent Publication No. 20110142917. In one embodiment the pharmaceutical composition is applied to the ear canal. Delivery to the ear may also be referred to as aural or otic delivery.

[0639] In one embodiment, the compositions, systems, or components thereof and/or vectors or vector systems can be delivered to ear via a transfection to the inner ear through the intact round window by a novel proteidic delivery technology which may be applied to the nucleic acid-targeting system of the present invention (see, e.g., Qi et al., Gene Therapy (2013), 1-9). About 40 µl of 10mM RNA may be contemplated as the dosage for administration to the ear.

[0640] According to Rejali et al. (Hear Res. 2007 Jun;228(1-2):180-7), cochlear implant function can be improved by good preservation of the spiral ganglion neurons, which are the target of electrical stimulation by the implant and brain derived neurotrophic factor (BDNF) has previously been shown to enhance spiral ganglion survival in experimentally deafened ears. Rejali et al. tested a modified design of the cochlear implant electrode that includes a coating of fibroblast cells transduced by a viral vector with a BDNF gene insert. To accomplish this type of *ex vivo* gene transfer, Rejali et al. transduced guinea pig fibroblasts with an adenovirus with a BDNF gene cassette insert and determined that these cells secreted BDNF and then attached BDNF-secreting cells to the cochlear implant electrode via an agarose gel, and implanted the electrode in the scala tympani. Rejali et al. determined that the BDNF expressing electrodes were able to preserve significantly more spiral ganglion neurons in the basal turns of the cochlea after 48 days of implantation when compared to control electrodes and demonstrated the feasibility of combining cochlear implant therapy with *ex vivo* gene transfer for enhancing spiral ganglion neuron survival. Such a system may be applied to the nucleic acid-targeting system of the present invention for delivery to the ear.

[0641] In one embodiment, the system set forth in Mukherjea et al. (Antioxidants & Redox Signaling, Volume 13, Number 5, 2010) can be adapted for transtympanic administration of the composition, system, or component thereof to the ear. In one embodiment, a dosage of about 2 mg to about 4 mg of chimeric IscB systems for administration to a human.

[0642] In one embodiment, the system set forth in [Jung et al. (Molecular Therapy, vol. 21 no. 4, 834–841 apr. 2013) can be adapted for vestibular epithelial delivery of the composition, system, or component thereof to the ear. In one embodiment, a dosage of about 1 to about 30 mg of chimeric IscB system for administration to a human.

Treating Diseases in Non-Dividing Cells

[0643] In one embodiment, the gene or transcript to be corrected is in a non-dividing cell. Exemplary non-dividing cells are muscle cells or neurons. Non-dividing (especially non-

dividing, fully differentiated) cell types present issues for gene targeting or genome engineering, for example because homologous recombination (HR) is generally suppressed in the G1 cell-cycle phase. However, while studying the mechanisms by which cells control normal DNA repair systems, Durocher discovered a previously unknown switch that keeps HR "off" in non-dividing cells and devised a strategy to toggle this switch back on. Orthwein et al. (Daniel Durocher's lab at the Mount Sinai Hospital in Ottawa, Canada) recently reported (Nature 16142, published online 9 Dec 2015) have shown that the suppression of HR can be lifted and gene targeting successfully concluded in both kidney (293T) and osteosarcoma (U2OS) cells. Tumor suppressors, BRCA1, PALB2 and BRAC2 are known to promote DNA DSB repair by HR. They found that formation of a complex of BRCA1 with PALB2 - BRAC2 is governed by a ubiquitin site on PALB2, such that action on the site by an E3 ubiquitin ligase. This E3 ubiquitin ligase is composed of KEAP1 (a PALB2 -interacting protein) in complex with cullin-3 (CUL3)–RBX1. PALB2 ubiquitylation suppresses its interaction with BRCA1 and is counteracted by the deubiquitylase USP11, which is itself under cell cycle control. Restoration of the BRCA1–PALB2 interaction combined with the activation of DNA-end resection is sufficient to induce homologous recombination in G1, as measured by a number of methods including a chimeric IscB system-based gene-targeting assay directed at USP11 or KEAP1 (expressed from a pX459 vector). However, when the BRCA1–PALB2 interaction was restored in resection-competent G1 cells using either KEAP1 depletion or expression of the PALB2-KR mutant, a robust increase in gene-targeting events was detected. These teachings can be adapted for and/or applied to the compositions, systems, described herein.

[0644] Thus, reactivation of HR in cells, especially non-dividing, fully differentiated cell types is preferred, In one embodiment. In one embodiment, promotion of the BRCA1–PALB2 interaction is preferred In one embodiment. In one embodiment, the target ell is a non-dividing cell. In one embodiment, the target cell is a neuron or muscle cell. In one embodiment, the target cell is targeted *in vivo*. In one embodiment, the cell is in G1 and HR is suppressed. In one embodiment, use of KEAP1 depletion, for example inhibition of expression of KEAP1 activity, is preferred. KEAP1 depletion may be achieved through siRNA, for example as shown in Orthwein et al. Alternatively, expression of the PALB2-KR mutant (lacking all eight Lys residues in the BRCA1-interaction domain is preferred, either in combination with KEAP1 depletion or alone. PALB2-KR interacts with BRCA1 irrespective of cell cycle position. Thus, promotion or restoration of the BRCA1-PALB2 interaction, especially in G1 cells, is preferred

In one embodiment, especially where the target cells are non-dividing, or where removal and return (*ex vivo* gene targeting) is problematic, for example neuron or muscle cells. KEAP1 siRNA is available from ThermoFischer. In one embodiment, a BRCA1–PALB2 complex may be delivered to the G1 cell. In one embodiment, PALB2 deubiquitylation may be promoted for example by increased expression of the deubiquitylase USP11, so it is envisaged that a construct may be provided to promote or up-regulate expression or activity of the deubiquitylase USP11.

Treating Diseases of the Eye

[0645] In one embodiment, the disease to be treated is a disease that affects the eyes. Thus, In one embodiment, the composition, system, or component thereof described herein is delivered to one or both eyes.

[0646] The composition, system, can be used to correct ocular defects that arise from several genetic mutations further described in Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012.

[0647] In one embodiment, the condition to be treated or targeted is an eye disorder. In one embodiment, the eye disorder may include glaucoma. In one embodiment, the eye disorder includes a retinal degenerative disease. In one embodiment, the retinal degenerative disease is selected from Stargardt disease, Bardet-Biedl Syndrome, Best disease, Blue Cone Monochromacy, Choroidermia, Cone-rod dystrophy, Congenital Stationary Night Blindness, Enhanced S-Cone Syndrome, Juvenile X-Linked Retinoschisis, Leber Congenital Amaurosis, Malattia Leventinesse, Norrie Disease or X-linked Familial Exudative Vitreoretinopathy, Pattern Dystrophy, Sorsby Dystrophy, Usher Syndrome, Retinitis Pigmentosa, Achromatopsia or Macular dystrophies or degeneration, Retinitis Pigmentosa, Achromatopsia, and age related macular degeneration. In one embodiment, the retinal degenerative disease is Leber Congenital Amaurosis (LCA) or Retinitis Pigmentosa. Other exemplary eye diseases are described in greater detail elsewhere herein.

[0648] In one embodiment, the composition, system, is delivered to the eye, optionally via intravitreal injection or subretinal injection. Intraocular injections may be performed with the aid of an operating microscope. For subretinal and intravitreal injections, eyes may be prolapsed by gentle digital pressure and fundi visualized using a contact lens system consisting of a drop of a coupling medium solution on the cornea covered with a glass microscope slide coverslip. For subretinal injections, the tip of a 10-mm 34-gauge needle, mounted on a 5-µl

Hamilton syringe may be advanced under direct visualization through the superior equatorial sclera tangentially towards the posterior pole until the aperture of the needle was visible in the subretinal space. Then, 2 µl of vector suspension may be injected to produce a superior bullous retinal detachment, thus confirming subretinal vector administration. This approach creates a self-sealing sclerotomy allowing the vector suspension to be retained in the subretinal space until it is absorbed by the RPE, usually within 48 h of the procedure. This procedure may be repeated in the inferior hemisphere to produce an inferior retinal detachment. This technique results in the exposure of approximately 70% of neurosensory retina and RPE to the vector suspension. For intravitreal injections, the needle tip may be advanced through the sclera 1 mm posterior to the corneoscleral limbus and 2 µl of vector suspension injected into the vitreous cavity. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 µl of vector suspension may be injected. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 µl of vector suspension may be injected. These vectors may be injected at titers of either 1.0–1.4 × 10^{10} or 1.0–1.4 × 10^{9} transducing units (TU)/ml.

[0649] In one embodiment, for administration to the eye, lentiviral vectors. In one embodiment, the lentiviral vector is an equine infectious anemia virus (EIAV) vector. Exemplary EIAV vectors for eye delivery are described in Balagaan, J Gene Med 2006; 8: 275 – 285, Published online 21 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845; Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012), which can be adapted for use with the composition, system, described herein. In one embodiment, the dosage can be 1.1 x 10^{5} transducing units per eye (TU/eye) in a total volume of 100 µl.

[0650] Other viral vectors can also be used for delivery to the eye, such as AAV vectors, such as those described in Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006), Millington-Ward et al. (Molecular Therapy, vol. 19 no. 4, 642–649 Apr. 2011; Dalkara et al. (Sci Transl Med 5, 189ra76 (2013)), which can be adapted for use with the composition, system, described herein. In one embodiment, the dose can range from about 10^{6} to $10^{9.5}$ particle units. In the context of the Millington-Ward AAV vectors, a dose of about 2 x 10^{11} to about 6 x 10^{13} virus particles can be administered. In the context of Dalkara vectors, a dose of about 1 x 10^{15} to about 1 x 10^{16} vg/ml administered to a human.

[0651] In one embodiment, the sd-rxRNA® system of RXi Pharmaceuticals may be used/and or adapted for delivering composition, system, to the eye. In this system, a single intravitreal administration of 3 μg of sd-rxRNA results in sequence-specific reduction of PPIB mRNA levels for 14 days. The sd-rxRNA® system may be applied to the nucleic acid-targeting system of the present invention, contemplating a dose of about 3 to 20 mg of composition administered to a human.

[0652] In other embodiments, the methods of US Patent Publication No. 20130183282, which is directed to methods of cleaving a target sequence from the human rhodopsin gene, may also be modified to the nucleic acid-targeting system of the present invention.

[0653] In other embodiments, the methods of US Patent Publication No. 20130202678 for treating retinopathies and sight-threatening ophthalmologic disorders relating to delivering of the Puf-A gene (which is expressed in retinal ganglion and pigmented cells of eye tissues and displays a unique anti-apoptotic activity) to the sub-retinal or intravitreal space in the eye may be used or adapted. In particular, desirable targets are zgc:193933, prdm1a, spata2, tex10, rbb4, ddx3, zp2.2, Blimp-1 and HtrA2, all of which may be targeted by the composition, system, of the present invention.

[0654] Wu (Cell Stem Cell,13:659–62, 2013) designed a guide RNA that led Cas9to a single base pair mutation that causes cataracts in mice, where it induced DNA cleavage. Then using either the other wild-type allele or oligos given to the zygotes repair mechanisms corrected the sequence of the broken allele and corrected the cataract-causing genetic defect in mutant mouse. This approach can be adapted to and/or applied to the compositions, systems, described herein.

[0655] US Patent Publication No. 20120159653, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with macular degeneration (MD), the teachings of which can be applied to and/or adapted for the compositions, systems, described herein.

[0656] One aspect of US Patent Publication No. 20120159653 relates to editing of any chromosomal sequences that encode proteins associated with MD which may be applied to the nucleic acid-targeting system of the present invention.

Treating Muscle Diseases and Cardiovascular Diseases

[0657] In one embodiment, the composition, system can be used to treat and/or prevent a muscle disease and associated circulatory or cardiovascular disease or disorder. The present

invention also contemplates delivering the composition, system, described herein, e.g., IscB polypeptide or chimeric IscB systems, to the heart. For the heart, a myocardium tropic adeno-associated virus (AAVM) is preferred, in particular AAVM41 which showed preferential gene transfer in the heart (see, e.g., Lin-Yanga et al., PNAS, March 10, 2009, vol. 106, no. 10). Administration may be systemic or local. A dosage of about 1-10 x 10^{14} vector genomes are contemplated for systemic administration. See also, e.g., Eulalio et al. (2012) Nature 492: 376 and Somasuntharam et al. (2013) Biomaterials 34: 7790, the teachings of which can be adapted for and/or applied to the compositions, systems, described herein.

[0658] For example, US Patent Publication No. 20110023139, the teachings of which can be adapted for and/or applied to the compositions, systems, described herein describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with cardiovascular disease. Cardiovascular diseases generally include high blood pressure, heart attacks, heart failure, and stroke and TIA. Any chromosomal sequence involved in cardiovascular disease, or the protein encoded by any chromosomal sequence involved in cardiovascular disease, may be utilized in the methods described in this disclosure. The cardiovascular-related proteins are typically selected based on an experimental association of the cardiovascular-related protein to the development of cardiovascular disease. For example, the production rate or circulating concentration of a cardiovascular-related protein may be elevated or depressed in a population having a cardiovascular disorder relative to a population lacking the cardiovascular disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the cardiovascular-related proteins may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[0659] The compositions, systems, herein can be used for treating diseases of the muscular system. The present invention also contemplates delivering the composition, system, described herein, effector protein systems, to muscle(s).

[0660] In one embodiment, the muscle disease to be treated is a muscle dystrophy such as DMD. In one embodiment, the composition, system, such as a system capable of RNA modification, described herein can be used to achieve exon skipping to achieve correction of

the diseased gene. As used herein, the term "exon skipping" refers to the modification of pre-mRNA splicing by the targeting of splice donor and/or acceptor sites within a pre-mRNA with one or more complementary antisense oligonucleotide(s) (AONs). By blocking access of a spliceosome to one or more splice donor or acceptor site, an AON may prevent a splicing reaction thereby causing the deletion of one or more exons from a fully-processed mRNA. Exon skipping may be achieved in the nucleus during the maturation process of pre-mRNAs. In some examples, exon skipping may include the masking of key sequences involved in the splicing of targeted exons by using a composition, system, described herein capable of RNA modification. In one embodiment, exon skipping can be achieved in dystrophin mRNA. In one embodiment, the composition, system, can induce exon skipping at exon 1, 2, 3, 4, 5, 6, 7, 8, 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, 45, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, or any combination thereof of the dystrophin mRNA. In one embodiment, the composition, system, can induce exon skipping at exon 43, 44, 50, 51, 52, 55, or any combination thereof of the dystrophin mRNA. Mutations in these exons, can also be corrected using non-exon skipping polynucleotide modification methods.

[0661] In one embodiment, for treatment of a muscle disease, the method of Bortolanza et al. Molecular Therapy vol. 19 no. 11, 2055–2064 Nov. 2011) may be applied to an AAV expressing chimeric IscB systems and injected into humans at a dosage of about 2×10^{15} or 2×10^{16} vg of vector. The teachings of Bortolanza et al., can be adapted for and/or applied to the compositions, systems, described herein.

[0662] In one embodiment, the method of Dumonceaux et al. (Molecular Therapy vol. 18 no. 5, 881–887 May 2010) may be applied to an AAV expressing chimeric IscB systems and injected into humans, for example, at a dosage of about 10^{14} to about 10^{15} vg of vector. The teachings of Dumonceaux described herein can be adapted for and/or applied to the compositions, systems, described herein.

[0663] In one embodiment, the method of Kinouchi et al. (Gene Therapy (2008) 15, 1126–1130) may be applied to compositions described herein and injected into a human, for example, at a dosage of about 500 to 1000 ml of a 40 μM solution into the muscle.

[0664] In one embodiment, the method of Hagstrom et al. (Molecular Therapy Vol. 10, No. 2, August 2004) can be adapted for and/or applied to the compositions, systems, herein and injected at a dose of about 15 to about 50 mg into the great saphenous vein of a human.

[0665] In one embodiment, the method comprises treating a sickle cell related disease, e.g., sickle cell trait, sickle cell disease such as sickle cell anemia, β-thalassaemia. For example, the method and system may be used to modify the genome of the sickle cell, e.g., by correcting one or more mutations of the β-globin gene. In the case of β-thalassaemia, sickle cell anemia can be corrected by modifying HSCs with the systems. The system allows the specific editing of the cell's genome by cutting its DNA and then letting it repair itself. The chimeric IscB system is inserted and directed by a RNA guide to the mutated point and then it cuts the DNA at that point. Simultaneously, a healthy version of the sequence is inserted. This sequence is used by the cell's own repair system to fix the induced cut. In this way, the chimeric IscB system allows the correction of the mutation in the previously obtained stem cells. The methods and systems may be used to correct HSCs as to sickle cell anemia using a systems that targets and corrects the mutation (e.g., with a suitable HDR template that delivers a coding sequence for β-globin, advantageously non-sickling β-globin); specifically, the guide RNA can target mutation that give rise to sickle cell anemia, and the HDR can provide coding for proper expression of β-globin. An ωRNA or guide RNA that targets the mutation-and- chimeric IscB system containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of β-globin; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier. The HDR template can provide for the HSC to express an engineered β-globin gene (e.g., βA-T87Q), or β-globin.

Treating Diseases of the Liver and Kidney

[0666] In one embodiment, the composition, system, or component thereof described herein can be used to treat a disease of the kidney or liver. Thus, in one embodiment, delivery of the composition or component thereof described herein is to the liver or kidney.

[0667] Delivery strategies to induce cellular uptake of the therapeutic nucleic acid include physical force or vector systems such as viral-, lipid- or complex- based delivery, or nanocarriers. From the initial applications with less possible clinical relevance, when nucleic acids were addressed to renal cells with hydrodynamic high-pressure injection systemically, a

wide range of gene therapeutic viral and non-viral carriers have been applied already to target posttranscriptional events in different animal kidney disease models *in vivo* (Csaba Révész and Péter Hamar (2011). Delivery Methods to Target RNAs in the Kidney, Gene Therapy Applications, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-541-9, InTech, Available from: www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-inthe-kidney). Delivery methods to the kidney may include those in Yuan et al. (Am J Physiol Renal Physiol 295: F605–F617, 2008). The method of Yuang et al. may be applied to the composition of the present invention contemplating a 1-2 g subcutaneous injection of chimeric IscB system conjugated with cholesterol to a human for delivery to the kidneys. In one embodiment, the method of Molitoris et al. (J Am Soc Nephrol 20: 1754–1764, 2009) can be adapted to the composition and a cumulative dose of 12- 20 mg/kg to a human can be used for delivery to the proximal tubule cells of the kidneys. In one embodiment, the methods of Thompson et al. (Nucleic Acid Therapeutics, Volume 22, Number 4, 2012) can be adapted to the compositions and a dose of up to 25 mg/kg can be delivered via i.v. administration. In one embodiment, the method of Shimizu et al. (J Am Soc Nephrol 21: 622–633, 2010) can be adapted to the compositions and a dose of about of 10-20 μmol compositions complexed with nanocarriers in about 1-2 liters of a physiologic fluid for i.p. administration can be used.

[0668] Other various delivery vehicles can be used to deliver the composition, system to the kidney such as viral, hydrodynamic, lipid, polymer nanoparticles, aptamers and various combinations thereof (see e.g. Larson et al., Surgery, (Aug 2007), Vol. 142, No. 2, pp. (262-269); Hamar et al., Proc Natl Acad Sci, (Oct 2004), Vol. 101, No. 41, pp. (14883-14888); Zheng et al., Am J Pathol, (Oct 2008), Vol. 173, No. 4, pp. (973–980); Feng et al., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283–1289); Q. Zhang et al., PloS ONE, (Jul 2010), Vol. 5, No. 7, e11709, pp. (1-13); Kushibikia et al., J Controlled Release, (Jul 2005), Vol. 105, No. 3, pp. (318-331); Wang et al., Gene Therapy, (Jul 2006), Vol. 13, No. 14, pp. (1097-1103); Kobayashi et al., Journal of Pharmacology and Experimental Therapeutics, (Feb 2004), Vol. 308, No. 2, pp. (688-693); Wolfrum et al., Nature Biotechnology, (Sep 2007), Vol. 25, No. 10, pp. (1149-1157); Molitoris et al., J Am Soc Nephrol, (Aug 2009), Vol. 20, No. 8 pp. (1754-1764); Mikhaylova et al., Cancer Gene Therapy, (Mar 2011), Vol. 16, No. 3, pp. (217-226); Y. Zhang et al., J Am Soc Nephrol, (Apr 2006), Vol. 17, No. 4, pp. (1090–1101); Singhal et al., Cancer Res, (May 2009), Vol. 69, No. 10, pp. (4244-4251); Malek et al., Toxicology and Applied Pharmacology, (Apr 2009), Vol. 236, No. 1, pp. (97-108); Shimizu et

al., J Am Soc Nephrology, (Apr 2010), Vol. 21, No. 4, pp. (622-633); Jiang et al., Molecular Pharmaceutics, (May-Jun 2009), Vol. 6, No. 3, pp. (727-737); Cao et al, J Controlled Release, (Jun 2010), Vol. 144, No. 2, pp. (203-212); Ninichuk et al., Am J Pathol, (Mar 2008), Vol. 172, No. 3, pp. (628-637); Purschke et al., Proc Natl Acad Sci, (Mar 2006), Vol. 103, No. 13, pp. (5173-5178).

[0669] In one embodiment, delivery is to liver cells. In one embodiment, the liver cell is a hepatocyte. Delivery of the composition and system herein may be via viral vectors, especially AAV (and in particular AAV2/6) vectors. These can be administered by intravenous injection. A preferred target for the liver, whether *in vitro* or *in vivo*, is the albumin gene. This is a so-called 'safe harbor" as albumin is expressed at very high levels and so some reduction in the production of albumin following successful gene editing is tolerated. It is also preferred as the high levels of expression seen from the albumin promoter/enhancer allows for useful levels of correct or transgene production (from the inserted recombination template) to be achieved even if only a small fraction of hepatocytes are edited. See sites identified by Wechsler et al. (reported at the 57th Annual Meeting and Exposition of the American Society of Hematology - abstract available online at ash.confex.com/ash/2015/webprogram/Paper86495.html and presented on 6th December 2015) which can be adapted for use with the compositions, systems, herein.

[0670] Exemplary liver and kidney diseases that can be treated and/or prevented are described elsewhere herein.

Treating Epithelial and Lung Diseases

[0671] In one embodiment, the disease treated or prevented by the composition and system described herein can be a lung or epithelial disease. The compositions and systems described herein can be used for treating epithelial and/or lung diseases. The present invention also contemplates delivering the composition, system, described herein, to one or both lungs.

[0672] In one embodiment, as viral vector can be used to deliver the composition, system, or component thereof to the lungs. In one embodiment, the AAV is an AAV-1, AAV-2, AAV-5, AAV-6, and/or AAV-9 for delivery to the lungs (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-2077 Dec 2009). In one embodiment, the MOI can vary from 1×10^3 to 4×10^5 vector genomes/cell. In one embodiment, the delivery vector can be an RSV vector as in Zamora et al. (Am J Respir Crit Care Med Vol 183. pp 531–538, 2011. The method of Zamora et al. may be applied to the nucleic acid-targeting system of the present invention and an

aerosolized composition, for example with a dosage of 0.6 mg/kg, may be contemplated for the present invention.

[0673] Subjects treated for a lung disease may for example receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. As such, aerosolized delivery is preferred for AAV delivery in general. An adenovirus or an AAV particle may be used for delivery. Suitable gene constructs, each operably linked to one or more regulatory sequences, may be cloned into the delivery vector. In this instance, the following constructs are provided as examples: Cbh or EF1a promoter for Cas, U6 or H1 promoter for guide RNA),: A preferred arrangement is to use a CFTRdelta508 targeting guide, a repair template for deltaF508 mutation and a codon optimized composition, with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs.

Treating Diseases of the Skin

[0674] The compositions and systems described herein can be used for the treatment of skin diseases. The present invention also contemplates delivering the composition and system, described herein, to the skin.

[0675] In one embodiment, delivery to the skin (intradermal delivery) of the composition, system, or component thereof can be via one or more microneedles or microneedle containing device. For example, In one embodiment the device and methods of Hickerson et al. (Molecular Therapy—Nucleic Acids (2013) 2, e129) can be used and/or adapted to deliver the composition, system, described herein, for example, at a dosage of up to 300 μl of 0.1 mg/ml compositions to the skin.

[0676] In one embodiment, the methods and techniques of Leachman et al. (Molecular Therapy, vol. 18 no. 2, 442–446 Feb. 2010) can be used and/or adapted for delivery of a compositions described herein to the skin.

[0677] In one embodiment, the methods and techniques of Zheng et al. (PNAS, July 24, 2012, vol. 109, no. 30, 11975–11980) can be used and/or adapted for nanoparticle delivery of a compositions described herein to the skin. In one embodiment, as dosage of about 25 nM applied in a single application can achieve gene knockdown in the skin.

Treating Cancer

[0678] The compositions, systems, described herein can be used for the treatment of cancer. The present invention also contemplates delivering the composition, system, described

herein, to a cancer cell. Also, as is described elsewhere herein the compositions, systems, can be used to modify an immune cell, such as a CAR or CAR T cell, which can then in turn be used to treat and/or prevent cancer. This is also described in International Patent Publication No. WO 2015/161276, the disclosure of which is hereby incorporated by reference and described herein below.

[0679] Target genes suitable for the treatment or prophylaxis of cancer can include those set forth in Tables 7 and 8. In one embodiment, target genes for cancer treatment and prevention can also include those described in International Patent Publication No. WO 2015/048577 the disclosure of which is hereby incorporated by reference and can be adapted for and/or applied to the composition, system, described herein.

Adoptive Cell Therapy

[0680] The compositions, systems, and components thereof described herein can be used to modify cells for an adoptive cell therapy. In an aspect of the invention, methods and compositions which involve editing a target nucleic acid sequence, or modulating expression of a target nucleic acid sequence, and applications thereof in connection with cancer immunotherapy are comprehended by adapting the composition, system, of the present invention. In some examples, the compositions, systems, and methods may be used to modify a stem cell (e.g., induced pluripotent cell) to derive modified natural killer cells, gamma delta T cells, and alpha beta T cells, which can be used for the adoptive cell therapy. In certain examples, the compositions, systems, and methods may be used to modify modified natural killer cells, gamma delta T cells, and alpha beta T cells.

[0681] As used herein, “ACT”, “adoptive cell therapy” and “adoptive cell transfer” may be used interchangeably. In an embodiment, Adoptive cell therapy (ACT) can refer to the transfer of cells to a patient with the goal of transferring the functionality and characteristics into the new host by engraftment of the cells (see, e.g., Mettananda et al., Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-thalassemia, Nat Commun. 2017 Sep 4;8(1):424). As used herein, the term "engraft" or "engraftment" refers to the process of cell incorporation into a tissue of interest *in vivo* through contact with existing cells of the tissue. Adoptive cell therapy (ACT) can refer to the transfer of cells, most commonly immune-derived cells, back into the same patient or into a new recipient host with the goal of transferring the immunologic functionality and characteristics into the new host. If possible, use of autologous cells helps the recipient by minimizing GVHD issues. The adoptive

transfer of autologous tumor infiltrating lymphocytes (TIL) (Zacharakis et al., (2018) Nat Med. 2018 Jun;24(6):724-730; Besser et al., (2010) Clin. Cancer Res 16 (9) 2646–55; Dudley et al., (2002) Science 298 (5594): 850–4; and Dudley et al., (2005) Journal of Clinical Oncology 23 (10): 2346–57.) or genetically re-directed peripheral blood mononuclear cells (Johnson et al., (2009) Blood 114 (3): 535–46; and Morgan et al., (2006) Science 314(5796) 126-9) has been used to successfully treat patients with advanced solid tumors, including melanoma, metastatic breast cancer and colorectal carcinoma, as well as patients with CD19-expressing hematologic malignancies (Kalos et al., (2011) Science Translational Medicine 3 (95): 95ra73). In an embodiment, allogenic cells immune cells are transferred (see, e.g., Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266). As described further herein, allogenic cells can be edited to reduce alloreactivity and prevent graft-versus-host disease. Thus, use of allogenic cells allows for cells to be obtained from healthy donors and prepared for use in patients as opposed to preparing autologous cells from a patient after diagnosis.

[0682] Aspects of the invention involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens or tumor specific neoantigens (see, e.g., Maus et al., 2014, Adoptive Immunotherapy for Cancer or Viruses, Annual Review of Immunology, Vol. 32: 189-225; Rosenberg and Restifo, 2015, Adoptive cell transfer as personalized immunotherapy for human cancer, *Science* Vol. 348 no. 6230 pp. 62-68; Restifo et al., 2015, Adoptive immunotherapy for cancer: harnessing the T cell response. *Nat. Rev. Immunol.* 12(4): 269-281; and Jenson and Riddell, 2014, Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells. Immunol Rev. 257(1): 127–144; and Rajasagi et al., 2014, Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood.* 2014 Jul 17;124(3):453-62).

[0683] In an embodiment, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: MR1 (see, e.g., Crowther, et al., 2020, Genome-wide CRISPR–Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1, *Nature Immunology* vol. 21, pages178–185), B cell maturation antigen (BCMA) (see, e.g., Friedman et al., Effective Targeting of Multiple BCMA-Expressing Hematological Malignancies by Anti-BCMA CAR T Cells, *Hum Gene Ther.* 2018 Mar 8; Berdeja JG, et al. Durable clinical responses in heavily pretreated patients with relapsed/refractory multiple myeloma: updated results from a

multicenter study of bb2121 anti-Bcma CAR T cell therapy. Blood. 2017;130:740; and Mouhieddine and Ghobrial, Immunotherapy in Multiple Myeloma: The Era of CAR T Cell Therapy, Hematologist, May-June 2018, Volume 15, issue 3); PSA (prostate-specific antigen); prostate-specific membrane antigen (PSMA); PSCA (Prostate stem cell antigen); Tyrosine-protein kinase transmembrane receptor ROR1; fibroblast activation protein (FAP); Tumor-associated glycoprotein 72 (TAG72); Carcinoembryonic antigen (CEA); Epithelial cell adhesion molecule (EPCAM); Mesothelin; Human Epidermal growth factor Receptor 2 (ERBB2 (Her2/neu)); Prostase; Prostatic acid phosphatase (PAP); elongation factor 2 mutant (ELF2M); Insulin-like growth factor 1 receptor (IGF-1R); gplOO; BCR-ABL (breakpoint cluster region-Abelson); tyrosinase; New York esophageal squamous cell carcinoma 1 (NY-ESO-1); κ-light chain, LAGE (L antigen); MAGE (melanoma antigen); Melanoma-associated antigen 1 (MAGE-A1); MAGE A3; MAGE A6; legumain; Human papillomavirus (HPV) E6; HPV E7; prostein; survivin; PCTA1 (Galectin 8); Melan-A/MART-1; Ras mutant; TRP-1 (tyrosinase related protein 1, or gp75); Tyrosinase-related Protein 2 (TRP2); TRP-2/INT2 (TRP-2/intron 2); RAGE (renal antigen); receptor for advanced glycation end products 1 (RAGE1); Renal ubiquitous 1, 2 (RU1, RU2); intestinal carboxyl esterase (iCE); Heat shock protein 70-2 (HSP70-2) mutant; thyroid stimulating hormone receptor (TSHR); CD123; CD171; CD19; CD20; CD22; CD26; CD30; CD33; CD44v7/8 (cluster of differentiation 44, exons 7/8); CD53; CD92; CD100; CD148; CD150; CD200; CD261; CD262; CD362; CS-1 (CD2 subset 1, CRACC, SLAMF7, CD319, and 19A24); C-type lectin-like molecule-1 (CLL-1); ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); Tn antigen (Tn Ag); Fms-Like Tyrosine Kinase 3 (FLT3); CD38; CD138; CD44v6; B7H3 (CD276); KIT (CD117); Interleukin-13 receptor subunit alpha-2 (IL-13Ra2); Interleukin 11 receptor alpha (IL-11Ra); prostate stem cell antigen (PSCA); Protease Serine 21 (PRSS21); vascular endothelial growth factor receptor 2 (VEGFR2); Lewis(Y) antigen; CD24; Platelet-derived growth factor receptor beta (PDGFR-beta); stage-specific embryonic antigen-4 (SSEA-4); Mucin 1, cell surface associated (MUC1); mucin 16 (MUC16); epidermal growth factor receptor (EGFR); epidermal growth factor receptor variant III (EGFRvIII); neural cell adhesion molecule (NCAM); carbonic anhydrase IX (CAIX); Proteasome (Prosome, Macropain) Subunit, Beta Type, 9 (LMP2); ephrin type-A receptor 2 (EphA2); Ephrin B2; Fucosyl GM1; sialyl Lewis adhesion molecule (sLe); ganglioside GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); TGS5; high molecular weight-melanoma-associated antigen

(HMWMAA); o-acetyl-GD2 ganglioside (OAcGD2); Folate receptor alpha; Folate receptor beta; tumor endothelial marker 1 (TEM1/CD248); tumor endothelial marker 7-related (TEM7R); claudin 6 (CLDN6); G protein-coupled receptor class C group 5, member D (GPRC5D); chromosome X open reading frame 61 (CXORF61); CD97; CD179a; anaplastic lymphoma kinase (ALK); Polysialic acid; placenta-specific 1 (PLAC1); hexasaccharide portion of globoH glycoceramide (GloboH); mammary gland differentiation antigen (NY-BR-1); uroplakin 2 (UPK2); Hepatitis A virus cellular receptor 1 (HAVCR1); adrenoceptor beta 3 (ADRB3); pannexin 3 (PANX3); G protein-coupled receptor 20 (GPR20); lymphocyte antigen 6 complex, locus K 9 (LY6K); Olfactory receptor 51E2 (OR51E2); TCR Gamma Alternate Reading Frame Protein (TARP); Wilms tumor protein (WT1); ETS translocation-variant gene 6, located on chromosome 12p (ETV6-AML); sperm protein 17 (SPA17); X Antigen Family, Member 1A (XAGE1); angiopoietin-binding cell surface receptor 2 (Tie 2); CT (cancer/testis (antigen)); melanoma cancer testis antigen-1 (MAD-CT-1); melanoma cancer testis antigen-2 (MAD-CT-2); Fos-related antigen 1; p53; p53 mutant; human Telomerase reverse transcriptase (hTERT); sarcoma translocation breakpoints; melanoma inhibitor of apoptosis (ML-IAP); ERG (transmembrane protease, serine 2 (TMPRSS2) ETS fusion gene); N-Acetyl glucosaminyl-transferase V (NA17); paired box protein Pax-3 (PAX3); Androgen receptor; Cyclin B1; Cyclin D1; v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Ras Homolog Family Member C (RhoC); Cytochrome P450 1B1 (CYP1B1); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS); Squamous Cell Carcinoma Antigen Recognized By T Cells-1 or 3 (SART1, SART3); Paired box protein Pax-5 (PAX5); proacrosin binding protein sp32 (OY-TES1); lymphocyte-specific protein tyrosine kinase (LCK); A kinase anchor protein 4 (AKAP-4); synovial sarcoma, X breakpoint-1, -2, -3 or -4 (SSX1, SSX2, SSX3, SSX4); CD79a; CD79b; CD72; Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Fc fragment of IgA receptor (FCAR); Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); CD300 molecule-like family member f (CD300LF); C-type lectin domain family 12 member A (CLEC12A); bone marrow stromal cell antigen 2 (BST2); EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2); lymphocyte antigen 75 (LY75); Glypican-3 (GPC3); Fc receptor-like 5 (FCRL5); mouse double minute 2 homolog (MDM2); livin; alphafetoprotein (AFP); transmembrane activator and CAML Interactor (TACI); B-cell activating factor receptor (BAFF-R); V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS); immunoglobulin lambda-like

polypeptide 1 (IGLL1); 707-AP (707 alanine proline); ART-4 (adenocarcinoma antigen recognized by T4 cells); BAGE (B antigen; b-catenin/m, b-catenin/mutated); CAMEL (CTL-recognized antigen on melanoma); CAP1 (carcinoembryonic antigen peptide 1); CASP-8 (caspase-8); CDC27m (cell-division cycle 27 mutated); CDK4/m (cycline-dependent kinase 4 mutated); Cyp-B (cyclophilin B); DAM (differentiation antigen melanoma); EGP-2 (epithelial glycoprotein 2); EGP-40 (epithelial glycoprotein 40); Erbb2, 3, 4 (erythroblastic leukemia viral oncogene homolog-2, -3, 4); FBP (folate binding protein); fAchR (Fetal acetylcholine receptor); G250 (glycoprotein 250); GAGE (G antigen); GnT-V (N-acetylglucosaminyltransferase V); HAGE (helicose antigen); ULA-A (human leukocyte antigen-A); HST2 (human signet ring tumor 2); KIAA0205; KDR (kinase insert domain receptor); LDLR/FUT (low density lipid receptor/GDP L-fucose: b-D-galactosidase 2-a-L fucosyltransferase); L1CAM (L1 cell adhesion molecule); MC1R (melanocortin 1 receptor); Myosin/m (myosin mutated); MUM-1, -2, -3 (melanoma ubiquitous mutated 1, 2, 3); NA88-A (NA cDNA clone of patient M88); KG2D (Natural killer group 2, member D) ligands; oncofetal antigen (h5T4); p190 minor bcr-abl (protein of 190KD bcr-abl); Pml/RARa (promyelocytic leukemia/retinoic acid receptor a); PRAME (preferentially expressed antigen of melanoma); SAGE (sarcoma antigen); TEL/AML1 (translocation Ets-family leukemia/acute myeloid leukemia 1); TPI/m (triosephosphate isomerase mutated); CD70; and any combination thereof.

[0684] In an embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-specific antigen (TSA).

[0685] In an embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a neoantigen.

[0686] In an embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-associated antigen (TAA).

[0687] In an embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a universal tumor antigen. In certain preferred embodiments, the universal tumor antigen is selected from the group consisting of: a human telomerase reverse transcriptase (hTERT),

survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B 1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (Dl), and any combinations thereof.

[0688] In an embodiment, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: CD19, BCMA, CD70, CLL-1, MAGE A3, MAGE A6, HPV E6, HPV E7, WT1, CD22, CD171, ROR1, MUC16, and SSX2. In certain preferred embodiments, the antigen may be CD19. For example, CD19 may be targeted in hematologic malignancies, such as in lymphomas, more particularly in B-cell lymphomas, such as without limitation in diffuse large B-cell lymphoma, primary mediastinal b-cell lymphoma, transformed follicular lymphoma, marginal zone lymphoma, mantle cell lymphoma, acute lymphoblastic leukemia including adult and pediatric ALL, non-Hodgkin lymphoma, indolent non-Hodgkin lymphoma, or chronic lymphocytic leukemia. For example, BCMA may be targeted in multiple myeloma or plasma cell leukemia (see, e.g., 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic Chimeric Antigen Receptor T Cells Targeting B Cell Maturation Antigen). For example, CLL1 may be targeted in acute myeloid leukemia. For example, MAGE A3, MAGE A6, SSX2, and/or KRAS may be targeted in solid tumors. For example, HPV E6 and/or HPV E7 may be targeted in cervical cancer or head and neck cancer. For example, WT1 may be targeted in acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), chronic myeloid leukemia (CML), non-small cell lung cancer, breast, pancreatic, ovarian or colorectal cancers, or mesothelioma. For example, CD22 may be targeted in B cell malignancies, including non-Hodgkin lymphoma, diffuse large B-cell lymphoma, or acute lymphoblastic leukemia. For example, CD171 may be targeted in neuroblastoma, glioblastoma, or lung, pancreatic, or ovarian cancers. For example, ROR1 may be targeted in ROR1+ malignancies, including non-small cell lung cancer, triple negative breast cancer, pancreatic cancer, prostate cancer, ALL, chronic lymphocytic leukemia, or mantle cell lymphoma. For example, MUC16 may be targeted in MUC16ecto+ epithelial ovarian, fallopian tube or primary peritoneal cancer. For example, CD70 may be targeted in both hematologic malignancies as well as in solid cancers such as renal cell carcinoma (RCC), gliomas (e.g., GBM), and head and neck cancers (HNSCC). CD70 is expressed in both hematologic malignancies as well as in solid cancers, while its expression in normal tissues is

restricted to a subset of lymphoid cell types (see, e.g., 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic CRISPR Engineered Anti-CD70 CAR-T Cells Demonstrate Potent Preclinical Activity Against Both Solid and Hematological Cancer Cells).

[0689] Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected peptide specificity (see U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[0690] As an alternative to, or addition to, TCR modifications, chimeric antigen receptors (CARs) may be used in order to generate immuno-responsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described (see U.S. Patent Nos. 5,843,728; 5,851,828; 5,912,170; 6,004,811; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,211,422; and, PCT Publication WO 9215322).

[0691] In general, CARs are comprised of an extracellular domain, a transmembrane domain, and an intracellular domain, wherein the extracellular domain comprises an antigen-binding domain that is specific for a predetermined target. While the antigen-binding domain of a CAR is often an antibody or antibody fragment (e.g., a single chain variable fragment, scFv), the binding domain is not particularly limited so long as it results in specific recognition of a target. For example, In one embodiment, the antigen-binding domain may comprise a receptor, such that the CAR is capable of binding to the ligand of the receptor. Alternatively, the antigen-binding domain may comprise a ligand, such that the CAR is capable of binding the endogenous receptor of that ligand.

[0692] The antigen-binding domain of a CAR is generally separated from the transmembrane domain by a hinge or spacer. The spacer is also not particularly limited, and it is designed to provide the CAR with flexibility. For example, a spacer domain may comprise a portion of a human Fc domain, including a portion of the CH3 domain, or the hinge region of any immunoglobulin, such as IgA, IgD, IgE, IgG, or IgM, or variants thereof. Furthermore, the hinge region may be modified so as to prevent off-target binding by FcRs or other potential interfering objects. For example, the hinge may comprise an IgG4 Fc domain with or without

a S228P, L235E, and/or N297Q mutation (according to Kabat numbering) in order to decrease binding to FcRs. Additional spacers/hinges include, but are not limited to, CD4, CD8, and CD28 hinge regions.

[0693] The transmembrane domain of a CAR may be derived either from a natural or from a synthetic source. Where the source is natural, the domain may be derived from any membrane bound or transmembrane protein. Transmembrane regions of particular use in this disclosure may be derived from CD8, CD28, CD3, CD45, CD4, CD5, CDS, CD9, CD 16, CD22, CD33, CD37, CD64, CD80, CD86, CD 134, CD137, CD 154, TCR. Alternatively, the transmembrane domain may be synthetic, in which case it will comprise predominantly hydrophobic residues such as leucine and valine. Preferably a triplet of phenylalanine, tryptophan and valine will be found at each end of a synthetic transmembrane domain. Optionally, a short oligo- or polypeptide linker, preferably between 2 and 10 amino acids in length may form the linkage between the transmembrane domain and the cytoplasmic signaling domain of the CAR. A glycine-serine doublet provides a particularly suitable linker.

[0694] Alternative CAR constructs may be characterized as belonging to successive generations. First-generation CARs typically consist of a single-chain variable fragment of an antibody specific for an antigen, for example comprising a VL linked to a VH of a specific antibody, linked by a flexible linker, for example by a CD8α hinge domain and a CD8α transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3ζ or FcRγ (scFv-CD3ζ or scFv-FcRγ; see U.S. Patent No. 7,741,465; U.S. Patent No. 5,912,172; U.S. Patent No. 5,906,936). Second-generation CARs incorporate the intracellular domains of one or more costimulatory molecules, such as CD28, OX40 (CD134), or 4-1BB (CD137) within the endodomain (for example scFv-CD28/OX40/4-1BB-CD3ζ; see U.S. Patent Nos. 8,911,993; 8,916,381; 8,975,071; 9,101,584; 9,102,760; 9,102,761). Third-generation CARs include a combination of costimulatory endodomains, such a CD3ζ-chain, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, CD2, CD7, LIGHT, LFA-1, NKG2C, B7-H3, CD30, CD40, PD-1, or CD28 signaling domains (for example scFv-CD28-4-1BB-CD3ζ or scFv-CD28-OX40-CD3ζ; see U.S. Patent No. 8,906,682; U.S. Patent No. 8,399,645; U.S. Pat. No. 5,686,281; PCT Publication No. WO 2014/134165; PCT Publication No. WO 2012/079000). In an embodiment, the primary signaling domain comprises a functional signaling domain of a protein selected from the group consisting of CD3 zeta, CD3 gamma, CD3 delta, CD3 epsilon, common FcR gamma (FCERIG), FcR beta (Fc Epsilon R1b),

CD79a, CD79b, Fc gamma RIIa, DAP10, and DAP12. In certain preferred embodiments, the primary signaling domain comprises a functional signaling domain of CD3ζ or FcRγ. In an embodiment, the one or more costimulatory signaling domains comprise a functional signaling domain of a protein selected, each independently, from the group consisting of: CD27, CD28, 4-1BB (CD137), OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83, CDS, ICAM-1, GITR, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRF1), CD160, CD19, CD4, CD8 alpha, CD8 beta, IL2R beta, IL2R gamma, IL7R alpha, ITGA4, VLA1, CD49a, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, CD11d, ITGAE, CD103, ITGAL, CD11a, LFA-1, ITGAM, CD11b, ITGAX, CD11c, ITGB1, CD29, ITGB2, CD18, ITGB7, TNFR2, TRANCE/RANKL, DNAM1 (CD226), SLAMF4 (CD244, 2B4), CD84, CD96 (Tactile), CEACAM1, CRTAM, Ly9 (CD229), CD160 (BY55), PSGL1, CD100 (SEMA4D), CD69, SLAMF6 (NTB-A, Lyl08), SLAM (SLAMF1, CD150, IPO-3), BLAME (SLAMF8), SELPLG (CD162), LTBR, LAT, GADS, SLP-76, PAG/Cbp, NKp44, NKp30, NKp46, and NKG2D. In an embodiment, the one or more costimulatory signaling domains comprise a functional signaling domain of a protein selected, each independently, from the group consisting of: 4-1BB, CD27, and CD28. In an embodiment, a chimeric antigen receptor may have the design as described in U.S. Patent No. 7,446,190, comprising an intracellular domain of CD3ζ chain (such as amino acid residues 52-163 of the human CD3 zeta chain, as shown in SEQ ID NO: 14 of US 7,446,190), a signaling region from CD28 and an antigen-binding element (or portion or domain; such as scFv). The CD28 portion, when between the zeta chain portion and the antigen-binding element, may suitably include the transmembrane and signaling domains of CD28 (such as amino acid residues 114-220 of SEQ ID NO: 10, full sequence shown in SEQ ID NO: 6 of US 7,446,190; these can include the following portion of CD28 as set forth in Genbank identifier NM_006139. Alternatively, when the zeta sequence lies between the CD28 sequence and the antigen-binding element, intracellular domain of CD28 can be used alone (such as amino sequence set forth in SEQ ID NO: 9 of US 7,446,190). Hence, an embodiment employs a CAR comprising (a) a zeta chain portion comprising the intracellular domain of human CD3ζ chain, (b) a costimulatory signaling region, and (c) an antigen-binding element (or portion or domain), wherein the costimulatory signaling region comprises the amino acid sequence encoded by SEQ ID NO: 6 of US 7,446,190.

[0695] Alternatively, co-stimulation may be orchestrated by expressing CARs in antigen-specific T cells, chosen so as to be activated and expanded following engagement of their native αβTCR, for example by antigen on professional antigen-presenting cells, with attendant co-stimulation. In addition, additional engineered receptors may be provided on the immuno-responsive cells, for example to improve targeting of a T-cell attack and/or minimize side effects.

[0696] By means of an example and without limitation, Kochenderfer et al., (2009) J Immunother. 32 (7): 689-702 described anti-CD19 chimeric antigen receptors (CAR). FMC63-28Z CAR contained a single chain variable region moiety (scFv) recognizing CD19 derived from the FMC63 mouse hybridoma (described in Nicholson et al., (1997) Molecular Immunology 34: 1157–1165), a portion of the human CD28 molecule, and the intracellular component of the human TCR-ζ molecule. FMC63-CD828BBZ CAR contained the FMC63 scFv, the hinge and transmembrane regions of the CD8 molecule, the cytoplasmic portions of CD28 and 4-1BB, and the cytoplasmic component of the TCR-ζ molecule. The exact sequence of the CD28 molecule included in the FMC63-28Z CAR corresponded to Genbank identifier NM_006139; the sequence included all amino acids starting with the amino acid sequence IEVMYPPPY (SEQ ID NO: 2058) and continuing all the way to the carboxy-terminus of the protein. To encode the anti-CD19 scFv component of the vector, the authors designed a DNA sequence which was based on a portion of a previously published CAR (Cooper et al., (2003) Blood 101: 1637–1644). This sequence encoded the following components in frame from the 5' end to the 3' end: an XhoI site, the human granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor α-chain signal sequence, the FMC63 light chain variable region (as in Nicholson et al., supra), a linker peptide (as in Cooper et al., supra), the FMC63 heavy chain variable region (as in Nicholson et al., supra), and a NotI site. A plasmid encoding this sequence was digested with XhoI and NotI. To form the MSGV-FMC63-28Z retroviral vector, the XhoI and NotI-digested fragment encoding the FMC63 scFv was ligated into a second XhoI and NotI-digested fragment that encoded the MSGV retroviral backbone (as in Hughes et al., (2005) Human Gene Therapy 16: 457–472) as well as part of the extracellular portion of human CD28, the entire transmembrane and cytoplasmic portion of human CD28, and the cytoplasmic portion of the human TCR-ζ molecule (as in Maher et al., 2002) Nature Biotechnology 20: 70–75). The FMC63-28Z CAR is included in the KTE-C19 (axicabtagene ciloleucel) anti-CD19 CAR-T therapy product in development by Kite Pharma, Inc. for the treatment of inter alia

patients with relapsed/refractory aggressive B-cell non-Hodgkin lymphoma (NHL). Accordingly, in an embodiment, cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may express the FMC63-28Z CAR as described by Kochenderfer et al. (supra). Hence, in an embodiment, cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element (or portion or domain; such as scFv) that specifically binds to an antigen, an intracellular signaling domain comprising an intracellular domain of a CD3ζ chain, and a costimulatory signaling region comprising a signaling domain of CD28. Preferably, the CD28 amino acid sequence is as set forth in Genbank identifier NM_006139 (sequence version 1, 2 or 3) starting with the amino acid sequence IEVMYPPPY and continuing all the way to the carboxy terminus of the protein. Preferably, the antigen is CD19, more preferably the antigen-binding element is an anti-CD19 scFv, even more preferably the anti-CD19 scFv as described by Kochenderfer et al. (supra).

[0697] Additional anti-CD19 CARs are further described in International Patent Publication No. WO 2015/187528. More particularly Example 1 and Table 1 of WO2015187528, incorporated by reference herein, demonstrate the generation of anti-CD19 CARs based on a fully human anti-CD19 monoclonal antibody (47G4, as described in US20100104509) and murine anti-CD19 monoclonal antibody (as described in Nicholson et al. and explained above). Various combinations of a signal sequence (human CD8-alpha or GM-CSF receptor), extracellular and transmembrane regions (human CD8-alpha) and intracellular T-cell signaling domains (CD28-CD3ζ; 4-1BB-CD3ζ; CD27-CD3ζ; CD28-CD27-CD3ζ, 4-1BB-CD27-CD3ζ; CD27-4-1BB-CD3ζ; CD28-CD27-FcεRI gamma chain; or CD28-FcεRI gamma chain) were disclosed. Hence, in an embodiment, cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element that specifically binds to an antigen, an extracellular and transmembrane region as set forth in Table 1 of WO2015187528 and an intracellular T-cell signaling domain as set forth in Table 1 of No. WO 2015/187528. Preferably, the antigen is CD19, more preferably the antigen-binding element is an anti-CD19 scFv, even more preferably the mouse or human anti-CD19 scFv as described in Example 1 of. WO 2015/187528. In an embodiment, the CAR comprises, consists essentially of or consists of an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:

10, SEQ ID NO: 11, SEQ ID NO: 12, or SEQ ID NO: 13 as set forth in Table 1 of WO2015187528.

[0698] By means of an example and without limitation, chimeric antigen receptor that recognizes the CD70 antigen is described in WO2012058460A2 (see also, Park et al., CD70 as a target for chimeric antigen receptor T cells in head and neck squamous cell carcinoma, Oral Oncol. 2018 Mar; 78:145-150; and Jin et al., CD70, a novel target of CAR T-cell therapy for gliomas, Neuro Oncol. 2018 Jan 10;20(1):55-65). CD70 is expressed by diffuse large B-cell and follicular lymphoma and also by the malignant cells of Hodgkins lymphoma, Waldenstrom's macroglobulinemia and multiple myeloma, and by HTLV-1- and EBV-associated malignancies. (Agathanggelou et al. Am.J.Pathol. 1995;147: 1152-1160; Hunter et al., Blood 2004; 104:4881. 26; Lens et al., J Immunol. 2005;174:6212-6219; Baba et al., J Virol. 2008;82:3843-3852.) In addition, CD70 is expressed by non-hematological malignancies such as renal cell carcinoma and glioblastoma. (Junker et al., J Urol. 2005;173:2150-2153; Chahlavi et al., Cancer Res 2005;65:5428-5438) Physiologically, CD70 expression is transient and restricted to a subset of highly activated T, B, and dendritic cells.

[0699] By means of an example and without limitation, chimeric antigen receptor that recognizes BCMA has been described (see, e.g., US20160046724A1; WO2016014789A2; WO2017211900A1; WO2015158671A1; US20180085444A1; WO2018028647A1; US20170283504A1; and WO2013154760A1).

[0700] In an embodiment, the immune cell may, in addition to a CAR or exogenous TCR as described herein, further comprise a chimeric inhibitory receptor (inhibitory CAR) that specifically binds to a second target antigen and is capable of inducing an inhibitory or immunosuppressive or repressive signal to the cell upon recognition of the second target antigen. In an embodiment, the chimeric inhibitory receptor comprises an extracellular antigen-binding element (or portion or domain) configured to specifically bind to a target antigen, a transmembrane domain, and an intracellular immunosuppressive or repressive signaling domain. In an embodiment, the second target antigen is an antigen that is not expressed on the surface of a cancer cell or infected cell or the expression of which is downregulated on a cancer cell or an infected cell. In an embodiment, the second target antigen is an MHC-class I molecule. In an embodiment, the intracellular signaling domain comprises a functional signaling portion of an immune checkpoint molecule, such as for example PD-1 or CTLA4.

Advantageously, the inclusion of such inhibitory CAR reduces the chance of the engineered immune cells attacking non-target (e.g., non-cancer) tissues.

[0701] Alternatively, T-cells expressing CARs may be further modified to reduce or eliminate expression of endogenous TCRs in order to reduce off-target effects. Reduction or elimination of endogenous TCRs can reduce off-target effects and increase the effectiveness of the T cells (U.S. 9,181,527). T cells stably lacking expression of a functional TCR may be produced using a variety of approaches. T cells internalize, sort, and degrade the entire T cell receptor as a complex, with a half-life of about 10 hours in resting T cells and 3 hours in stimulated T cells (von Essen, M. et al. 2004. J. Immunol. 173:384-393). Proper functioning of the TCR complex requires the proper stoichiometric ratio of the proteins that compose the TCR complex. TCR function also requires two functioning TCR zeta proteins with ITAM motifs. The activation of the TCR upon engagement of its MHC-peptide ligand requires the engagement of several TCRs on the same T cell, which all must signal properly. Thus, if a TCR complex is destabilized with proteins that do not associate properly or cannot signal optimally, the T cell will not become activated sufficiently to begin a cellular response.

[0702] Accordingly, in one embodiment, TCR expression may eliminated using RNA interference (e.g., shRNA, siRNA, miRNA, etc.), chimeric IscB systems, or other methods that target the nucleic acids encoding specific TCRs (e.g., TCR-α and TCR-β) and/or CD3 chains in primary T cells. By blocking expression of one or more of these proteins, the T cell will no longer produce one or more of the key components of the TCR complex, thereby destabilizing the TCR complex and preventing cell surface expression of a functional TCR.

[0703] In some instances, CAR may also comprise a switch mechanism for controlling expression and/or activation of the CAR. For example, a CAR may comprise an extracellular, transmembrane, and intracellular domain, in which the extracellular domain comprises a target-specific binding element that comprises a label, binding domain, or tag that is specific for a molecule other than the target antigen that is expressed on or by a target cell. In such embodiments, the specificity of the CAR is provided by a second construct that comprises a target antigen binding domain (e.g., an scFv or a bispecific antibody that is specific for both the target antigen and the label or tag on the CAR) and a domain that is recognized by or binds to the label, binding domain, or tag on the CAR. See, e.g., WO 2013/044225, WO 2016/000304, WO 2015/057834, WO 2015/057852, WO 2016/070061, US 9,233,125, US 2016/0129109. In this way, a T-cell that expresses the CAR can be administered to a subject,

but the CAR cannot bind its target antigen until the second composition comprising an antigen-specific binding domain is administered.

[0704] Alternative switch mechanisms include CARs that require multimerization in order to activate their signaling function (see, e.g., US Patent Publication Nos. US 2015/0368342, US 2016/0175359, US 2015/0368360) and/or an exogenous signal, such as a small molecule drug (US 2016/0166613, Yung et al., Science, 2015), in order to elicit a T-cell response. Some CARs may also comprise a "suicide switch" to induce cell death of the CAR T-cells following treatment (Buddee et al., PLoS One, 2013) or to downregulate expression of the CAR following binding to the target antigen (International Patent Publication No. WO 2016/011210).

[0705] Alternative techniques may be used to transform target immuno-responsive cells, such as protoplast fusion, lipofection, transfection or electroporation. A wide variety of vectors may be used, such as retroviral vectors, lentiviral vectors, adenoviral vectors, adeno-associated viral vectors, plasmids or transposons, such as a Sleeping Beauty transposon (see U.S. Patent Nos. 6,489,458; 7,148,203; 7,160,682; 7,985,739; 8,227,432), may be used to introduce CARs, for example using 2nd generation antigen-specific CARs signaling through CD3ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[0706] Cells that are targeted for transformation may for example include T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated. T cells expressing a desired CAR may for example be selected through co-culture with γ-irradiated activating and propagating cells (AaPC), which co-express the cancer antigen and co-stimulatory molecules. The engineered CAR T-cells may be expanded, for example by co-culture on AaPC in presence of soluble factors, such as IL-2 and IL-21. This expansion may for example be carried out so as to provide memory CAR+ T cells (which may for example be assayed by non-enzymatic digital array and/or multi-panel flow cytometry). In this way, CAR T cells may be provided that have specific cytotoxic activity against antigen-bearing tumors (optionally in conjunction with production of desired chemokines such as interferon-γ). CAR T cells of this kind may for example be used in animal models, for example to treat tumor xenografts.

[0707] In an embodiment, ACT includes co-transferring CD4+ Th1 cells and CD8+ CTLs to induce a synergistic antitumor response (see, e.g., Li et al., Adoptive cell therapy with CD4+

T helper 1 cells and CD8+ cytotoxic T cells enhances complete rejection of an established tumor, leading to generation of endogenous memory responses to non-targeted tumor epitopes. Clin Transl Immunology. 2017 Oct; 6(10): e160).

[0708] In an embodiment, Th17 cells are transferred to a subject in need thereof. Th17 cells have been reported to directly eradicate melanoma tumors in mice to a greater extent than Th1 cells (Muranski P, et al., Tumor-specific Th17-polarized cells eradicate large established melanoma. Blood. 2008 Jul 15; 112(2):362-73; and Martin-Orozco N, et al., T helper 17 cells promote cytotoxic T cell activation in tumor immunity. Immunity. 2009 Nov 20; 31(5):787-98). Those studies involved an adoptive T cell transfer (ACT) therapy approach, which takes advantage of CD4+ T cells that express a TCR recognizing tyrosinase tumor antigen. Exploitation of the TCR leads to rapid expansion of Th17 populations to large numbers *ex vivo* for reinfusion into the autologous tumor-bearing hosts.

[0709] In an embodiment, ACT may include autologous iPSC-based vaccines, such as irradiated iPSCs in autologous anti-tumor vaccines (see e.g., Kooreman, Nigel G. et al., Autologous iPSC-Based Vaccines Elicit Anti-tumor Responses *In Vivo*, Cell Stem Cell 22, 1–13, 2018, doi.org/10.1016/j.stem.2018.01.016).

[0710] Unlike T-cell receptors (TCRs) that are MHC restricted, CARs can potentially bind any cell surface-expressed antigen and can thus be more universally used to treat patients (see Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, Front. Immunol., 03 April 2017, doi.org/10.3389/fimmu.2017.00267). In an embodiment, in the absence of endogenous T-cell infiltrate (e.g., due to aberrant antigen processing and presentation), which precludes the use of TIL therapy and immune checkpoint blockade, the transfer of CAR T-cells may be used to treat patients (see, e.g., Hinrichs CS, Rosenberg SA. Exploiting the curative potential of adoptive T-cell therapy for cancer. Immunol Rev (2014) 257(1):56–71. doi:10.1111/ imr.12132).

[0711] Approaches such as the foregoing may be adapted to provide methods of treating and/or increasing survival of a subject having a disease, such as a neoplasia, for example by administering an effective amount of an immuno-responsive cell comprising an antigen recognizing receptor that binds a selected antigen, wherein the binding activates the immuno-responsive cell, thereby treating or preventing the disease (such as a neoplasia, a pathogen infection, an autoimmune disorder, or an allogeneic transplant reaction).

[0712] In an embodiment, the treatment can be administered after lymphodepleting pretreatment in the form of chemotherapy (typically a combination of cyclophosphamide and fludarabine) or radiation therapy. Initial studies in ACT had short lived responses and the transferred cells did not persist *in vivo* for very long (Houot et al., T-cell-based immunotherapy: adoptive cell transfer and checkpoint inhibition. Cancer Immunol Res (2015) 3(10):1115–22; and Kamta et al., Advancing Cancer Therapy with Present and Emerging Immuno-Oncology Approaches. Front. Oncol. (2017) 7:64). Immune suppressor cells like Tregs and MDSCs may attenuate the activity of transferred cells by outcompeting them for the necessary cytokines. Not being bound by a theory lymphodepleting pretreatment may eliminate the suppressor cells allowing the TILs to persist.

[0713] In one embodiment, the treatment can be administrated into patients undergoing an immunosuppressive treatment (e.g., glucocorticoid treatment). The cells, or population of cells, may be made resistant to at least one immunosuppressive agent due to the inactivation of a gene encoding a receptor for such immunosuppressive agent. In an embodiment, the immunosuppressive treatment provides for the selection and expansion of the immuno-responsive T cells within the patient.

[0714] In an embodiment, the treatment can be administered before primary treatment (e.g., surgery or radiation therapy) to shrink a tumor before the primary treatment. In another embodiment, the treatment can be administered after primary treatment to remove any remaining cancer cells.

[0715] In an embodiment, immuno-metabolic barriers can be targeted therapeutically prior to and/or during ACT to enhance responses to ACT or CAR T-cell therapy and to support endogenous immunity (see, e.g., Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, Front. Immunol., 03 April 2017, doi.org/10.3389/fimmu.2017.00267).

[0716] The administration of cells or population of cells, such as immune system cells or cell populations, such as more particularly immuno-responsive cells or cell populations, as disclosed herein may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The cells or population of cells may be administered to a patient subcutaneously, intradermally, intratumorally, intranodally, intramedullary, intramuscularly, intrathecally, by intravenous or intralymphatic injection, or intraperitoneally. In one embodiment, the disclosed CARs may be delivered or

administered into a cavity formed by the resection of tumor tissue (i.e. intracavity delivery) or directly into a tumor prior to resection (i.e. intratumoral delivery). In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

[0717] The administration of the cells or population of cells can consist of the administration of 104- 109 cells per kg body weight, preferably 105 to 106 cells/kg body weight including all integer values of cell numbers within those ranges. Dosing in CAR T cell therapies may for example involve administration of from 106 to 109 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide. The cells or population of cells can be administrated in one or more doses. In another embodiment, the effective amount of cells are administrated as a single dose. In another embodiment, the effective amount of cells are administrated as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a blood bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions are within the skill of one in the art. An effective amount means an amount which provides a therapeutic or prophylactic benefit. The dosage administrated will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

[0718] In another embodiment, the effective amount of cells or composition comprising those cells are administrated parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[0719] To guard against possible adverse reactions, engineered immuno-responsive cells may be equipped with a transgenic safety switch, in the form of a transgene that renders the cells vulnerable to exposure to a specific signal. For example, the herpes simplex viral thymidine kinase (TK) gene may be used in this way, for example by introduction into allogeneic T lymphocytes used as donor lymphocyte infusions following stem cell transplantation (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. Front. Pharmacol. 2015; 6: 95). In such cells, administration of a nucleoside prodrug such as ganciclovir or acyclovir causes cell death. Alternative safety switch constructs include inducible caspase 9, for example triggered by administration of a small-molecule dimerizer that brings together two nonfunctional icasp9 molecules to form the active enzyme. A wide variety of alternative approaches to implementing cellular proliferation controls have been

described (see U.S. Patent Publication No. 20130071414; International Patent Publication WO 2011/146862; International Patent Publication WO 2014/011987; International Patent Publication WO 2013/040371; Zhou et al. BLOOD, 2014, 123/25:3895 – 3905; Di Stasi et al., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos et al., Stem Cells 28(6):1107-15 (2010)).

[0720] In a further refinement of adoptive therapies, genome editing may be used to tailor immuno-responsive cells to alternative implementations, for example providing edited CAR T cells (see Poirot et al., 2015, Multiplex genome edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853; Ren et al., 2017, Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition, Clin Cancer Res. 2017 May 1;23(9):2255-2266. doi: 10.1158/1078-0432.CCR-16-1300. Epub 2016 Nov 4; Qasim et al., 2017, Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells, Sci Transl Med. 2017 Jan 25;9(374); Legut, et al., 2018, CRISPR-mediated TCR replacement generates superior anticancer transgenic T cells. Blood, 131(3), 311-322; and Georgiadis et al., Long Terminal Repeat CRISPR-CAR-Coupled "Universal" T Cells Mediate Potent Anti-leukemic Effects, Molecular Therapy, In Press, Corrected Proof, Available online 6 March 2018). Cells may be edited using any CRISPR system and method of use thereof as described herein. The composition and systems may be delivered to an immune cell by any method described herein. In preferred embodiments, cells are edited *ex vivo* and transferred to a subject in need thereof. Immuno-responsive cells, CAR T cells or any cells used for adoptive cell transfer may be edited. Editing may be performed for example to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell (e.g. TRAC locus); to eliminate potential alloreactive T-cell receptors (TCR) or to prevent inappropriate pairing between endogenous and exogenous TCR chains, such as to knock-out or knock-down expression of an endogenous TCR in a cell; to disrupt the target of a chemotherapeutic agent in a cell; to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell; to knock-out or knock-down expression of other gene or genes in a cell, the reduced expression or lack of expression of which can enhance the efficacy of adoptive therapies using the cell; to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene encoding an antigen targeted by an exogenous CAR or TCR; to knock-out or knock-down expression of one or more MHC constituent proteins in a cell; to activate a T cell;

to modulate cells such that the cells are resistant to exhaustion or dysfunction; and/or increase the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8+ T-cells (see International Patent Publication Nos. WO 2013/176915, WO 2014/059173, WO 2014/172606, WO 2014/184744, and WO 2014/191128).

[0721] In an embodiment, editing may result in inactivation of a gene. By inactivating a gene, it is intended that the gene of interest is not expressed in a functional protein form. In a particular embodiment, the system specifically catalyzes cleavage in one targeted gene thereby inactivating said targeted gene. The nucleic acid strand breaks caused are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). However, NHEJ is an imperfect repair process that often results in changes to the DNA sequence at the site of the cleavage. Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions (Indel) and can be used for the creation of specific gene knockouts. Cells in which a cleavage induced mutagenesis event has occurred can be identified and/or selected by well-known methods in the art. In an embodiment, homology directed repair (HDR) is used to concurrently inactivate a gene (e.g., TRAC) and insert an endogenous TCR or CAR into the inactivated locus.

[0722] Hence, in an embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may be performed to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell. Conventionally, nucleic acid molecules encoding CARs or TCRs are transfected or transduced to cells using randomly integrating vectors, which, depending on the site of integration, may lead to clonal expansion, oncogenic transformation, variegated transgene expression and/or transcriptional silencing of the transgene. Directing of transgene(s) to a specific locus in a cell can minimize or avoid such risks and advantageously provide for uniform expression of the transgene(s) by the cells. Without limitation, suitable 'safe harbor' loci for directed transgene integration include CCR5 or AAVS1. Homology-directed repair (HDR) strategies are known and described elsewhere in this specification allowing to insert transgenes into desired loci (e.g., TRAC locus).

[0723] Further suitable loci for insertion of transgenes, in particular CAR or exogenous TCR transgenes, include without limitation loci comprising genes coding for constituents of endogenous T-cell receptor, such as T-cell receptor alpha locus (TRA) or T-cell receptor beta locus (TRB), for example T-cell receptor alpha constant (TRAC) locus, T-cell receptor beta

constant 1 (TRBC1) locus or T-cell receptor beta constant 2 (TRBC1) locus. Advantageously, insertion of a transgene into such locus can simultaneously achieve expression of the transgene, potentially controlled by the endogenous promoter, and knock-out expression of the endogenous TCR. This approach has been exemplified in Eyquem et al., (2017) Nature 543: 113-117, wherein the authors used CRISPR/Cas9 gene editing to knock-in a DNA molecule encoding a CD19-specific CAR into the TRAC locus downstream of the endogenous promoter; the CAR-T cells obtained by CRISPR were significantly superior in terms of reduced tonic CAR signaling and exhaustion.

[0724] T cell receptors (TCR) are cell surface receptors that participate in the activation of T cells in response to the presentation of antigen. The TCR is generally made from two chains, α and β, which assemble to form a heterodimer and associates with the CD3-transducing subunits to form the T cell receptor complex present on the cell surface. Each α and β chain of the TCR consists of an immunoglobulin-like N-terminal variable (V) and constant (C) region, a hydrophobic transmembrane domain, and a short cytoplasmic region. As for immunoglobulin molecules, the variable region of the α and β chains are generated by V(D)J recombination, creating a large diversity of antigen specificities within the population of T cells. However, in contrast to immunoglobulins that recognize intact antigen, T cells are activated by processed peptide fragments in association with an MHC molecule, introducing an extra dimension to antigen recognition by T cells, known as MHC restriction. Recognition of MHC disparities between the donor and recipient through the T cell receptor leads to T cell proliferation and the potential development of graft versus host disease (GVHD). The inactivation of TCRα or TCRβ can result in the elimination of the TCR from the surface of T cells preventing recognition of alloantigen and thus GVHD. However, TCR disruption generally results in the elimination of the CD3 signaling component and alters the means of further T cell expansion.

[0725] Hence, in an embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may be performed to knock-out or knock-down expression of an endogenous TCR in a cell. For example, NHEJ-based or HDR-based gene editing approaches can be employed to disrupt the endogenous TCR alpha and/or beta chain genes. For example, gene editing system or systems, such as IscB system or systems, can be designed to target a sequence found within the TCR beta chain conserved between the beta 1 and beta 2 constant region genes (TRBC1 and TRBC2) and/or to target the constant region of the TCR alpha chain (TRAC) gene.

[0726] Allogeneic cells are rapidly rejected by the host immune system. It has been demonstrated that, allogeneic leukocytes present in non-irradiated blood products will persist for no more than 5 to 6 days (Boni, Muranski et al. 2008 Blood 1;112(12):4746-54). Thus, to prevent rejection of allogeneic cells, the host's immune system usually has to be suppressed to some extent. However, in the case of adoptive cell transfer the use of immunosuppressive drugs also have a detrimental effect on the introduced therapeutic T cells. Therefore, to effectively use an adoptive immunotherapy approach in these conditions, the introduced cells would need to be resistant to the immunosuppressive treatment. Thus, in a particular embodiment, the present invention further comprises a step of modifying T cells to make them resistant to an immunosuppressive agent, preferably by inactivating at least one gene encoding a target for an immunosuppressive agent. An immunosuppressive agent is an agent that suppresses immune function by one of several mechanisms of action. An immunosuppressive agent can be, but is not limited to a calcineurin inhibitor, a target of rapamycin, an interleukin-2 receptor α-chain blocker, an inhibitor of inosine monophosphate dehydrogenase, an inhibitor of dihydrofolic acid reductase, a corticosteroid or an immunosuppressive antimetabolite. The present invention allows conferring immunosuppressive resistance to T cells for immunotherapy by inactivating the target of the immunosuppressive agent in T cells. As non-limiting examples, targets for an immunosuppressive agent can be a receptor for an immunosuppressive agent such as: CD52, glucocorticoid receptor (GR), a FKBP family gene member and a cyclophilin family gene member.

[0727] In an embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may be performed to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell. Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In an embodiment, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[0728] Additional immune checkpoints include Src homology 2 domain-containing protein tyrosine phosphatase 1 (SHP-1) (Watson HA, et al., SHP-1: the next checkpoint target for cancer immunotherapy? Biochem Soc Trans. 2016 Apr 15;44(2):356-62). SHP-1 is a widely expressed inhibitory protein tyrosine phosphatase (PTP). In T-cells, it is a negative regulator of antigen-dependent activation and proliferation. It is a cytosolic protein, and therefore not amenable to antibody-mediated therapies, but its role in activation and proliferation makes it an attractive target for genetic manipulation in adoptive transfer strategies, such as chimeric antigen receptor (CAR) T cells. Immune checkpoints may also include T cell immunoreceptor with Ig and ITIM domains (TIGIT/Vstm3/WUCAM/VSIG9) and VISTA (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. Front. Immunol. 6:418).

[0729] International Patent Publication No. WO 2014/172606 relates to the use of MT1 and/or MT2 inhibitors to increase proliferation and/or activity of exhausted CD8+ T-cells and to decrease CD8+ T-cell exhaustion (e.g., decrease functionally exhausted or unresponsive CD8+ immune cells). In an embodiment, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[0730] In an embodiment, targets of gene editing may be at least one targeted locus involved in the expression of an immune checkpoint protein. Such targets may include, but are not limited to CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBRII, TGFRBRI, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1, TIM-3, CEACAM-1, CEACAM-3, or CEACAM-5. In preferred embodiments, the gene locus involved in the expression of PD-1 or CTLA-4 genes is targeted. In other preferred embodiments, combinations of genes are targeted, such as but not limited to PD-1 and TIGIT.

[0731] By means of an example and without limitation, International Patent Publication No. WO 2016/196388 concerns an engineered T cell comprising (a) a genetically engineered antigen receptor that specifically binds to an antigen, which receptor may be a CAR; and (b) a disrupted gene encoding a PD-L1, an agent for disruption of a gene encoding a PD- L1, and/or

disruption of a gene encoding PD-L1, wherein the disruption of the gene may be mediated by a gene editing nuclease, a zinc finger nuclease (ZFN), CRISPR/Cas9 and/or TALEN. WO2015142675 relates to immune effector cells comprising a CAR in combination with an agent (such as the composition or system herein) that increases the efficacy of the immune effector cells in the treatment of cancer, wherein the agent may inhibit an immune inhibitory molecule, such as PD1, PD-L1, CTLA-4, TIM-3, LAG-3, VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, TGFR beta, CEACAM-1, CEACAM-3, or CEACAM-5. Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas9 mRNA and gRNAs targeting endogenous TCR, β-2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1.

[0732] In an embodiment, cells may be engineered to express a CAR, wherein expression and/or function of methylcytosine dioxygenase genes (TET1, TET2 and/or TET3) in the cells has been reduced or eliminated, (such as the composition or system herein) (for example, as described in WO201704916).

[0733] In an embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may be performed to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene encoding an antigen targeted by an exogenous CAR or TCR, thereby reducing the likelihood of targeting of the engineered cells. In an embodiment, the targeted antigen may be one or more antigen selected from the group consisting of CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, CD362, human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (D1), B cell maturation antigen (BCMA), transmembrane activator and CAML Interactor (TACI), and B-cell activating factor receptor (BAFF-R) (for example, as described in International Patent Publication Nos. WO 2016/011210 and WO 2017/011804).

[0734] In an embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immuno-responsive cells such as T cells, may be performed to knock-out or knock-down expression of one or more MHC constituent proteins, such as one or

more HLA proteins and/or beta-2 microglobulin (B2M), in a cell, whereby rejection of non-autologous (e.g., allogeneic) cells by the recipient's immune system can be reduced or avoided. In preferred embodiments, one or more HLA class I proteins, such as HLA-A, B and/or C, and/or B2M may be knocked-out or knocked-down. Preferably, B2M may be knocked-out or knocked-down. By means of an example, Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas mRNA and gRNAs targeting endogenous TCR, β-2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1.

[0735] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCRα, PD1 and TCRβ, CTLA-4 and TCRα, CTLA-4 and TCRβ, LAG3 and TCRα, LAG3 and TCRβ, Tim3 and TCRα, Tim3 and TCRβ, BTLA and TCRα, BTLA and TCRβ, BY55 and TCRα, BY55 and TCRβ, TIGIT and TCRα, TIGIT and TCRβ, B7H5 and TCRα, B7H5 and TCRβ, LAIR1 and TCRα, LAIR1 and TCRβ, SIGLEC10 and TCRα, SIGLEC10 and TCRβ, 2B4 and TCRα, 2B4 and TCRβ, B2M and TCRα, B2M and TCRβ.

[0736] In an embodiment, a cell may be multiplied edited (multiplex genome editing) as taught herein to (1) knock-out or knock-down expression of an endogenous TCR (for example, TRBC1, TRBC2 and/or TRAC), (2) knock-out or knock-down expression of an immune checkpoint protein or receptor (for example PD1, PD-L1 and/or CTLA4); and (3) knock-out or knock-down expression of one or more MHC constituent proteins (for example, HLA-A, B and/or C, and/or B2M, preferably B2M).

[0737] Whether prior to or after genetic modification of the T cells, the T cells can be activated and expanded generally using methods as described, for example, in U.S. Patent Nos. 6,352,694; 6,534,055; 6,905,680; 5,858,358; 6,887,466; 6,905,681; 7,144,575; 7,232,566; 7,175,843; 5,883,223; 6,905,874; 6,797,514; 6,867,041; and 7,572,631. T cells can be expanded *in vitro* or *in vivo*.

[0738] Immune cells may be obtained using any method known in the art. In one embodiment, allogenic T cells may be obtained from healthy subjects. In one embodiment T cells that have infiltrated a tumor are isolated. T cells may be removed during surgery. T cells may be isolated after removal of tumor tissue by biopsy. T cells may be isolated by any means known in the art. In one embodiment, T cells are obtained by apheresis. In one embodiment, the method may comprise obtaining a bulk population of T cells from a tumor sample by any

suitable method known in the art. For example, a bulk population of T cells can be obtained from a tumor sample by dissociating the tumor sample into a cell suspension from which specific cell populations can be selected. Suitable methods of obtaining a bulk population of T cells may include, but are not limited to, any one or more of mechanically dissociating (e.g., mincing) the tumor, enzymatically dissociating (e.g., digesting) the tumor, and aspiration (e.g., as with a needle).

[0739] The bulk population of T cells obtained from a tumor sample may comprise any suitable type of T cell. Preferably, the bulk population of T cells obtained from a tumor sample comprises tumor infiltrating lymphocytes (TILs).

[0740] The tumor sample may be obtained from any mammal. Unless stated otherwise, as used herein, the term "mammal" refers to any mammal including, but not limited to, mammals of the order Logomorpha, such as rabbits; the order Carnivora, including Felines (cats) and Canines (dogs); the order Artiodactyla, including Bovines (cows) and Swines (pigs); or of the order Perssodactyla, including Equines (horses). The mammals may be non-human primates, e.g., of the order Primates, Ceboids, or Simoids (monkeys) or of the order Anthropoids (humans and apes). In one embodiment, the mammal may be a mammal of the order Rodentia, such as mice and hamsters. Preferably, the mammal is a non-human primate or a human. An especially preferred mammal is the human.

[0741] T cells can be obtained from a number of sources, including peripheral blood mononuclear cells (PBMC), bone marrow, lymph node tissue, spleen tissue, and tumors. In an embodiment of the present invention, T cells can be obtained from a unit of blood collected from a subject using any number of techniques known to the skilled artisan, such as Ficoll separation. In one preferred embodiment, cells from the circulating blood of an individual are obtained by apheresis or leukapheresis. The apheresis product typically contains lymphocytes, including T cells, monocytes, granulocytes, B cells, other nucleated white blood cells, red blood cells, and platelets. In one embodiment, the cells collected by apheresis may be washed to remove the plasma fraction and to place the cells in an appropriate buffer or media for subsequent processing steps. In one embodiment of the invention, the cells are washed with phosphate buffered saline (PBS). In an alternative embodiment, the wash solution lacks calcium and may lack magnesium or may lack many if not all divalent cations. Initial activation steps in the absence of calcium lead to magnified activation. As those of ordinary skill in the art would readily appreciate a washing step may be accomplished by methods known to those

in the art, such as by using a semi-automated "flow-through" centrifuge (for example, the Cobe 2991 cell processor) according to the manufacturer's instructions. After washing, the cells may be resuspended in a variety of biocompatible buffers, such as, for example, Ca-free, Mg-free PBS. Alternatively, the undesirable components of the apheresis sample may be removed, and the cells directly resuspended in culture media.

[0742] In another embodiment, T cells are isolated from peripheral blood lymphocytes by lysing the red blood cells and depleting the monocytes, for example, by centrifugation through a PERCOLL™ gradient. A specific subpopulation of T cells, such as CD28+, CD4+, CDC, CD45RA+, and CD45RO+ T cells, can be further isolated by positive or negative selection techniques. For example, in one preferred embodiment, T cells are isolated by incubation with anti-CD3/anti-CD28 (i.e., 3×28)-conjugated beads, such as DYNABEADS® M-450 CD3/CD28 T, or XCYTE DYNABEADS™ for a time period sufficient for positive selection of the desired T cells. In one embodiment, the time period is about 30 minutes. In a further embodiment, the time period ranges from 30 minutes to 36 hours or longer and all integer values there between. In a further embodiment, the time period is at least 1, 2, 3, 4, 5, or 6 hours. In yet another preferred embodiment, the time period is 10 to 24 hours. In one preferred embodiment, the incubation time period is 24 hours. For isolation of T cells from patients with leukemia, use of longer incubation times, such as 24 hours, can increase cell yield. Longer incubation times may be used to isolate T cells in any situation where there are few T cells as compared to other cell types, such in isolating tumor infiltrating lymphocytes (TIL) from tumor tissue or from immunocompromised individuals. Further, use of longer incubation times can increase the efficiency of capture of CD8+ T cells.

[0743] Enrichment of a T cell population by negative selection can be accomplished with a combination of antibodies directed to surface markers unique to the negatively selected cells. A preferred method is cell sorting and/or selection via negative magnetic immunoadherence or flow cytometry that uses a cocktail of monoclonal antibodies directed to cell surface markers present on the cells negatively selected. For example, to enrich for CD4+ cells by negative selection, a monoclonal antibody cocktail typically includes antibodies to CD14, CD20, CD11b, CD16, HLA-DR, and CD8.

[0744] Further, monocyte populations (e.g., CD14+ cells) may be depleted from blood preparations by a variety of methodologies, including anti-CD14 coated beads or columns, or utilization of the phagocytotic activity of these cells to facilitate removal. Accordingly, in one

embodiment, the invention uses paramagnetic particles of a size sufficient to be engulfed by phagocytotic monocytes. In an embodiment, the paramagnetic particles are commercially available beads, for example, those produced by Life Technologies under the trade name Dynabeads™. In one embodiment, other non-specific cells are removed by coating the paramagnetic particles with "irrelevant" proteins (e.g., serum proteins or antibodies). Irrelevant proteins and antibodies include those proteins and antibodies or fragments thereof that do not specifically target the T cells to be isolated. In an embodiment, the irrelevant beads include beads coated with sheep anti-mouse antibodies, goat anti-mouse antibodies, and human serum albumin.

[0745] In brief, such depletion of monocytes is performed by preincubating T cells isolated from whole blood, apheresed peripheral blood, or tumors with one or more varieties of irrelevant or non-antibody coupled paramagnetic particles at any amount that allows for removal of monocytes (approximately a 20:1 bead:cell ratio) for about 30 minutes to 2 hours at 22 to 37 degrees C., followed by magnetic removal of cells which have attached to or engulfed the paramagnetic particles. Such separation can be performed using standard methods available in the art. For example, any magnetic separation methodology may be used including a variety of which are commercially available, (e.g., DYNAL® Magnetic Particle Concentrator (DYNAL MPC®)). Assurance of requisite depletion can be monitored by a variety of methodologies known to those of ordinary skill in the art, including flow cytometric analysis of CD14 positive cells, before and after depletion.

[0746] For isolation of a desired population of cells by positive or negative selection, the concentration of cells and surface (e.g., particles such as beads) can be varied. In an embodiment, it may be desirable to significantly decrease the volume in which beads and cells are mixed together (i.e., increase the concentration of cells), to ensure maximum contact of cells and beads. For example, in one embodiment, a concentration of 2 billion cells/ml is used. In one embodiment, a concentration of 1 billion cells/ml is used. In a further embodiment, greater than 100 million cells/ml is used. In a further embodiment, a concentration of cells of 10, 15, 20, 25, 30, 35, 40, 45, or 50 million cells/ml is used. In yet another embodiment, a concentration of cells from 75, 80, 85, 90, 95, or 100 million cells/ml is used. In further embodiments, concentrations of 125 or 150 million cells/ml can be used. Using high concentrations can result in increased cell yield, cell activation, and cell expansion. Further, use of high cell concentrations allows more efficient capture of cells that may weakly express

target antigens of interest, such as CD28-negative T cells, or from samples where there are many tumor cells present (i.e., leukemic blood, tumor tissue, etc). Such populations of cells may have therapeutic value and would be desirable to obtain. For example, using high concentration of cells allows more efficient selection of CD8+ T cells that normally have weaker CD28 expression.

[0747] In a related embodiment, it may be desirable to use lower concentrations of cells. By significantly diluting the mixture of T cells and surface (e.g., particles such as beads), interactions between the particles and cells are minimized. This selects for cells that express high amounts of desired antigens to be bound to the particles. For example, CD4+ T cells express higher levels of CD28 and are more efficiently captured than CD8+ T cells in dilute concentrations. In one embodiment, the concentration of cells used is 5×106/ml. In other embodiments, the concentration used can be from about 1×105/ml to 1×106/ml, and any integer value in between.

[0748] T cells can also be frozen. Wishing not to be bound by theory, the freeze and subsequent thaw step provides a more uniform product by removing granulocytes and to some extent monocytes in the cell population. After a washing step to remove plasma and platelets, the cells may be suspended in a freezing solution. While many freezing solutions and parameters are known in the art and will be useful in this context, one method involves using PBS containing 20% DMSO and 8% human serum albumin, or other suitable cell freezing media, the cells then are frozen to −80° C at a rate of 1° per minute and stored in the vapor phase of a liquid nitrogen storage tank. Other methods of controlled freezing may be used as well as uncontrolled freezing immediately at −20° C. or in liquid nitrogen.

[0749] T cells for use in the present invention may also be antigen-specific T cells. For example, tumor-specific T cells can be used. In an embodiment, antigen-specific T cells can be isolated from a patient of interest, such as a patient afflicted with a cancer or an infectious disease. In one embodiment, neoepitopes are determined for a subject, and T cells specific to these antigens are isolated. Antigen-specific cells for use in expansion may also be generated *in vitro* using any number of methods known in the art, for example, as described in U.S. Patent Publication No. US 20040224402 entitled, Generation and Isolation of Antigen-Specific T Cells, or in U.S. Pat. No. 6,040,177. Antigen-specific cells for use in the present invention may also be generated using any number of methods known in the art, for example, as described in

Current Protocols in Immunology, or Current Protocols in Cell Biology, both published by John Wiley & Sons, Inc., Boston, Mass.

[0750] In a related embodiment, it may be desirable to sort or otherwise positively select (e.g., via magnetic selection) the antigen specific cells prior to or following one or two rounds of expansion. Sorting or positively selecting antigen-specific cells can be carried out using peptide-MHC tetramers (Altman, et al., Science. 1996 Oct. 4; 274(5284):94-6). In another embodiment, the adaptable tetramer technology approach is used (Andersen et al., 2012 Nat Protoc. 7:891-902). Tetramers are limited by the need to utilize predicted binding peptides based on prior hypotheses, and the restriction to specific HLAs. Peptide-MHC tetramers can be generated using techniques known in the art and can be made with any MHC molecule of interest and any antigen of interest as described herein. Specific epitopes to be used in this context can be identified using numerous assays known in the art. For example, the ability of a polypeptide to bind to MHC class I may be evaluated indirectly by monitoring the ability to promote incorporation of 125I labeled β2-microglobulin (β2m) into MHC class I/β2m/peptide heterotrimeric complexes (see Parker et al., J. Immunol. 152:163, 1994).

[0751] In one embodiment cells are directly labeled with an epitope-specific reagent for isolation by flow cytometry followed by characterization of phenotype and TCRs. In one embodiment, T cells are isolated by contacting with T cell specific antibodies. Sorting of antigen-specific T cells, or generally any cells of the present invention, can be carried out using any of a variety of commercially available cell sorters, including, but not limited to, MoFlo sorter (DakoCytomation, Fort Collins, Colo.), FACSAria™, FACSArray™, FACSVantage™, BD™ LSR II, and FACSCalibur™ (BD Biosciences, San Jose, Calif.).

[0752] In a preferred embodiment, the method comprises selecting cells that also express CD3. The method may comprise specifically selecting the cells in any suitable manner. Preferably, the selecting is carried out using flow cytometry. The flow cytometry may be carried out using any suitable method known in the art. The flow cytometry may employ any suitable antibodies and stains. Preferably, the antibody is chosen such that it specifically recognizes and binds to the particular biomarker being selected. For example, the specific selection of CD3, CD8, TIM-3, LAG-3, 4-1BB, or PD-1 may be carried out using anti-CD3, anti-CD8, anti-TIM-3, anti-LAG-3, anti-4-lBB, or anti-PD-1 antibodies, respectively. The antibody or antibodies may be conjugated to a bead (e.g., a magnetic bead) or to a fluorochrome. Preferably, the flow cytometry is fluorescence-activated cell sorting (FACS).

TCRs expressed on T cells can be selected based on reactivity to autologous tumors. Additionally, T cells that are reactive to tumors can be selected for based on markers using the methods described in patent publication Nos. WO2014133567 and WO2014133568, herein incorporated by reference in their entirety. Additionally, activated T cells can be selected for based on surface expression of CD107a.

[0753] In one embodiment of the invention, the method further comprises expanding the numbers of T cells in the enriched cell population. Such methods are described in U.S. Patent No. 8,637,307 and is herein incorporated by reference in its entirety. The numbers of T cells may be increased at least about 3-fold (or 4-, 5-, 6-, 7-, 8-, or 9-fold), more preferably at least about 10-fold (or 20-, 30-, 40-, 50-, 60-, 70-, 80-, or 90-fold), more preferably at least about 100-fold, more preferably at least about 1,000 fold, or most preferably at least about 100,000-fold. The numbers of T cells may be expanded using any suitable method known in the art. Exemplary methods of expanding the numbers of cells are described in patent publication No. WO 2003/057171, U.S. Patent No. 8,034,334, and U.S. Patent Publication No. 2012/0244133, each of which is incorporated herein by reference.

[0754] In one embodiment, *ex vivo* T cell expansion can be performed by isolation of T cells and subsequent stimulation or activation followed by further expansion. In one embodiment of the invention, the T cells may be stimulated or activated by a single agent. In another embodiment, T cells are stimulated or activated with two agents, one that induces a primary signal and a second that is a co-stimulatory signal. Ligands useful for stimulating a single signal or stimulating a primary signal and an accessory molecule that stimulates a second signal may be used in soluble form. Ligands may be attached to the surface of a cell, to an Engineered Multivalent Signaling Platform (EMSP), or immobilized on a surface. In a preferred embodiment both primary and secondary agents are co-immobilized on a surface, for example a bead or a cell. In one embodiment, the molecule providing the primary activation signal may be a CD3 ligand, and the co-stimulatory molecule may be a CD28 ligand or 4-1BB ligand.

[0755] In an embodiment, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in International Patent Publication No. WO 2015/120096, by a method comprising enriching a population of lymphocytes obtained from a donor subject; stimulating the population of lymphocytes with one or more T-cell stimulating agents to produce a population of activated T cells, wherein the stimulation is performed in a closed

system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using a single cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells for a predetermined time to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. In an embodiment, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in WO 2015/120096, by a method comprising: obtaining a population of lymphocytes; stimulating the population of lymphocytes with one or more stimulating agents to produce a population of activated T cells, wherein the stimulation is performed in a closed system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using at least one cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. The predetermined time for expanding the population of transduced T cells may be 3 days. The time from enriching the population of lymphocytes to producing the engineered T cells may be 6 days. The closed system may be a closed bag system. Further provided is population of T cells comprising a CAR or an exogenous TCR obtainable or obtained by said method, and a pharmaceutical composition comprising such cells.

[0756] In an embodiment, T cell maturation or differentiation *in vitro* may be delayed or inhibited by the method as described in International Patent Publication No. WO 2017/070395, comprising contacting one or more T cells from a subject in need of a T cell therapy with an AKT inhibitor (such as, e.g., one or a combination of two or more AKT inhibitors disclosed in claim 8 of WO2017070395) and at least one of exogenous Interleukin-7 (IL-7) and exogenous Interleukin-15 (IL-15), wherein the resulting T cells exhibit delayed maturation or differentiation, and/or wherein the resulting T cells exhibit improved T cell function (such as, e.g., increased T cell proliferation; increased cytokine production; and/or increased cytolytic activity) relative to a T cell function of a T cell cultured in the absence of an AKT inhibitor.

[0757] In an embodiment, a patient in need of a T cell therapy may be conditioned by a method as described in International Patent Publication No. WO 2016/191756 comprising

administering to the patient a dose of cyclophosphamide between 200 mg/m2/day and 2000 mg/m2/day and a dose of fludarabine between 20 mg/m2/day and 900 mg/m^2/day.

Diseases

Genetic Diseases and Diseases with a Genetic and/or Epigenetic Aspect

[0758] The compositions, systems, or components thereof can be used to treat and/or prevent a genetic disease or a disease with a genetic and/or epigenetic aspect. The genes and conditions exemplified herein are not exhaustive. In one embodiment, a method of treating and/or preventing a genetic disease can include administering a composition, system, and/or one or more components thereof to a subject, where the composition, system, and/or one or more components thereof is capable of modifying one or more copies of one or more genes associated with the genetic disease or a disease with a genetic and/or epigenetic aspect in one or more cells of the subject. In one embodiment, modifying one or more copies of one or more genes associated with a genetic disease or a disease with a genetic and/or epigenetic aspect in the subject can eliminate a genetic disease or a symptom thereof in the subject. In one embodiment, modifying one or more copies of one or more genes associated with a genetic disease or a disease with a genetic and/or epigenetic aspect in the subject can decrease the severity of a genetic disease or a symptom thereof in the subject. In one embodiment, the compositions, systems, or components thereof can modify one or more genes or polynucleotides associated with one or more diseases, including genetic diseases and/or those having a genetic aspect and/or epigenetic aspect, including but not limited to, any one or more set forth in Table 6. It will be appreciated that those diseases and associated genes listed herein are non-exhaustive and non-limiting. Further some genes play roles in the development of multiple diseases.

Table 6. Exemplary Genetic and Other Diseases and Associated Genes

Disease Name	Primary Tissues or System Affected	Additional Tissues/Systems Affected	Genes
Achondroplasia	Bone and Muscle		fibroblast growth factor receptor 3 (FGFR3)
Achromatopsia	eye		CNGA3, CNGB3, GNAT2, PDE6C, PDE6H, ACHM2, ACHM3,
Acute Renal Injury	kidney		NFkappaB, AATF, p85alpha, FAS, Apoptosis cascade elements (e.g. FASR, Caspase 2, 3, 4, 6, 7, 8, 9, 10, AKT, TNF alpha, IGF1, IGF1R, RIPK1), p53

Age Related Macular Degeneration	eye		Abcr; CCL2; CC2; CP (ceruloplasmin); Timp3; cathepsinD; VLDLR, CCR2
AIDS	Immune System		KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1
Albinism (including oculocutaneous albinism (types 1-7) and ocular albinism)	Skin, hair, eyes,		TYR, OCA2, TYRP1, and SLC45A2, SLC24A5 and C10orf11
Alkaptonuria	Metabolism of amino acids	Tissues/organs where homogentisic acid accumulates, particularly cartilage (joints), heart valves, kidneys	HGD
alpha-1 antitrypsin deficiency (AATD or A1AD)	Lung	Liver, skin, vascular system, kidneys, GI	SERPINA1, those set forth in WO2017165862, PiZ allele
ALS	CNS		SOD1; ALS2; ALS3; ALS5; ALS7;STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c); DPP6; NEFH, PTGS1, SLC1A2, TNFRSF10B, PRPH, HSP90AA1, CRIA2, IFNG, AMPA2 S100B, FGF2, AOX1, CS, TXN, RAPHJ1, MAP3K5, NBEAL1, GPX1, ICA1L, RAC1, MAPT, ITPR2, ALS2CR4, GLS, ALS2CR8, CNTFR, ALS2CR11, FOLH1, FAM117B, P4HB, CNTF, SQSTM1, STRADB, NAIP, NLR, YWHAQ, SLC33A1, TRAK2, SCA1, NIF3L1, NIF3, PARD3B, COX8A, CDK15, HECW1, HECT, C2, WW 15, NOS1, MET, SOD2, HSPB1, NEFL, CTSB, ANG, HSPA8, RNase A, VAPB, VAMP, SNCA, alpha HGF, CAT, ACTB, NEFM, TH, BCL2, FAS, CASP3, CLU, SMN1, G6PD, BAX, HSF1, RNF19A, JUN, ALS2CR12, HSPA5, MAPK14, APEX1, TXNRD1, NOS2, TIMP1, CASP9, XIAP, GLG1, EPO, VEGFA, ELN, GDNF, NFE2L2, SLC6A3, HSPA4, APOE, PSMB8, DCTN2, TIMP3, KIFAP3, SLC1A1, SMN2, CCNC, STUB1, ALS2, PRDX6, SYP, CABIN1, CASP1, GART, CDK5, ATXN3, RTN4, C1QB, VEGFC, HTT, PARK7, XDH, GFAP, MAP2, CYCS, FCGR3B, CCS, UBL5, MMP9m SLC18A3, TRPM7, HSPB2, AKT1, DEERL1, CCL2, NGRN, GSR, TPPP3, APAF1, BTBD10, GLUD1, CXCR4, S:C1A3, FLT1, PON1, AR, LIF, ERBB3,

			:GA:S1, CD44, TP53, TLR3, GRIA1, GAPDH, AMPA, GRIK1, DES, CHAT, FLT4, CHMP2B, BAG1, CHRNA4, GSS, BAK1, KDR, GSTP1, OGG1, IL6
Alzheimer's Disease	Brain		E1; CHIP; UCH; UBB; Tau; LRP; PICALM; CLU; PS1; SORL1; CR1; VLDLR; UBA1; UBA3; CHIP28; AQP1; UCHL1; UCHL3; APP, AAA, CVAP, AD1, APOE, AD2, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3, ALAS2, ABCA1, BIN1, BDNF, BTNL8, C1ORF49, CDH4, CHRNB2, CKLFSF2, CLEC4E, CR1L, CSF3R, CST3, CYP2C, DAPK1, ESR1, FCAR, FCGR3B, FFA2, FGA, GAB2, GALP, GAPDHS, GMPB, HP, HTR7, IDE, IF127, IFI6, IFIT2, IL1RN, IL-1RA, IL8RA, IL8RB, JAG1, KCNJ15, LRP6, MAPT, MARK4, MPHOSPH1, MTHFR, NBN, NCSTN, NIACR2, NMNAT3, NTM, ORM1, P2RY13, PBEF1, PCK1, PICALM, PLAU, PLXNC1, PRNP, PSEN1, PSEN2, PTPRA, RALGPS2, RGSL2, SELENBP1, SLC25A37, SORL1, Mitoferrin-1, TF, TFAM, TNF, TNFRSF10C, UBE1C
Amyloidosis			APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB
Amyloid neuropathy			TTR, PALB
Anemia	Blood		CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT
Angelman Syndrome	Nervous system, brain		UBE3A
Attention Deficit Hyperactivity Disorder (ADHD)	Brain		PTCHD1
Autoimmune lymphoproliferative syndrome	Immune system		TNFRSF6, APT1, FAS, CD95, ALPS1A
Autism, Autism spectrum disorders (ASDs), including Asperger's and a general diagnostic category called Pervasive Developmental Disorders (PDDs)	Brain		PTCHD1; Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; GLO1, RTT, PPMX, MRX16, RX79, NLGN3, NLGN4, KIAA1260, AUTSX2, FMR1, FMR2; FXR1; FXR2; MGLUR5, ATP10C, CDH10, GRM6, MGLUR6, CDH9, CNTN4, NLGN2, CNTNAP2, SEMA5A, DHCR7, NLGN4X, NLGN4Y, DPP6, NLGN5, EN2, NRCAM, MDGA2, NRXN1, FMR2, AFF2, FOXP2, OR4M2, OXTR, FXR1, FXR2, PAH,

			GABRA1, PTEN, GABRA5, PTPRZ1, GABRB3, GABRG1, HIRIP3, SEZ6L2, HOXA1, SHANK3, IL6, SHBZRAP1, LAMB1, SLC6A4, SERT, MAPK3, TAS2R1, MAZ, TSC1, MDGA2, TSC2, MECP2, UBE3A, WNT2, see also 20110023145
autosomal dominant polycystic kidney disease (ADPKD) - (includes diseases such as von Hippel-Lindau disease and tubreous sclerosis complex disease)	kidney	liver	PKD1, PKD2
Autosomal Recessive Polycystic Kidney Disease (ARPKD)	kidney	liver	PKDH1
Ataxia-Telangiectasia (a.k.a Louis Bar syndrome)	Nervous system, immune system	various	ATM
B-Cell Non-Hodgkin Lymphoma			BCL7A, BCL7
Bardet-Biedl syndrome	Eye, musculoskeletal system, kidney, reproductive organs	Liver, ear, gastrointestinal system, brain	ARL6, BBS1, BBS2, BBS4, BBS5, BBS7, BBS9, BBS10, BBS12, CEP290, INPP5E, LZTFL1, MKKS, MKS1, SDCCAG8, TRIM32, TTC8
Bare Lymphocyte Syndrome	blood		TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5
Bartter's Syndrome (types I, II, III, IVA and B, and V)	kidney		SLC12A1 (type I), KCNJ1 (type II), CLCNKB (type III), BSND (type IV A), or both the CLCNKA CLCNKB genes (type IV B), CASR (type V).
Becker muscular dystrophy	Muscle		DMD, BMD, MYF6
Best Disease (Vitelliform Macular Dystrophy type 2)	eye		VMD2
Bleeding Disorders	blood		TBXA2R, P2RX1, P2X1
Blue Cone Monochromacy	eye		OPN1LW, OPN1MW, and LCR
Breast Cancer	Breast tissue		BRCA1, BRCA2, COX-2
Bruton's Disease (aka X-linked Agammglobulinemia)	Immune system, specifically B cells		BTK
Cancers (e.g., lymphoma, chronic lymphocytic leukemia (CLL), B cell acute lymphocytic leukemia (B-ALL), acute lymphoblastic leukemia, acute myeloid leukemia, non-Hodgkin's lymphoma (NHL), diffuse large cell lymphoma (DLCL), multiple myeloma, renal cell carcinoma (RCC), neuroblastoma, colorectal cancer, breast cancer, ovarian cancer, melanoma, sarcoma, prostate cancer, lung cancer, esophageal cancer, hepatocellular carcinoma,	Various		FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC, TRBC, those described in WO2015048577

pancreatic cancer, astrocytoma, mesothelioma, head and neck cancer, and medulloblastoma			
Cardiovascular Diseases	heart	Vascular system	IL1B, XDH, TP53, PTGS, MB, IL4, ANGPT1, ABCGu8, CTSK, PTGIR, KCNJ11, INS, CRP, PDGFRB, CCNA2, PDGFB, KCNJ5, KCNN3, CAPN10, ADRA2B, ABCG5, PRDX2, CPAN5, PARP14, MEX3C, ACE, RNF, IL6, TNF, STN, SERPINE1, ALB, ADIPOQ, APOB, APOE, LEP, MTHFR, APOA1, EDN1, NPPB, NOS3, PPARG, PLAT, PTGS2, CETP, AGTR1, HMGCR, IGF1, SELE, REN, PPARA, PON1, KNG1, CCL2, LPL, VWF, F2, ICAM1, TGFB, NPPA, IL10, EPO, SOD1, VCAM1, IFNG, LPA, MPO, ESR1, MAPK, HP, F3, CST3, COG2, MMP9, SERPINC1, F8, HMOX1, APOC3, IL8, PROL1, CBS, NOS2, TLR4, SELP, ABCA1, AGT, LDLR, GPT, VEGFA, NR3C2, IL18, NOS1, NR3C1, FGB, HGF, IL1A, AKT1, LIPC, HSPD1, MAPK14, SPP1, ITGB3, CAT, UTS2, THBD, F10, CP, TNFRSF11B, EGFR, MMP2, PLG, NPY, RHOD, MAPK8, MYC, FN1, CMA1, PLAU, GNB3, ADRB2, SOD2, F5, VDR, ALOX5, HLA-DRB1, PARP1, CD40LG, PON2, AGER, IRS1, PTGS1, ECE1, F7, IRMN, EPHX2, IGFBP1, MAPK10, FAS, ABCB1, JUN, IGFBP3, CD14, PDE5A, AGTR2, CD40, LCAT, CCR5, MMP1, TIMP1, ADM, DYT10, STAT3, MMP3, ELN, USF1, CFH, HSPA4, MMP12, MME, F2R, SELL, CTSB, ANXA5, ADRB1, CYBA, FGA, GGT1, LIPG, HIF1A, CXCR4, PROC, SCARB1, CD79A, PLTP, ADD1, FGG, SAA1, KCNH2, DPP4, NPR1, VTN, KIAA0101, FOS, TLR2, PPIG, IL1R1, AR, CYP1A1, SERPINA1, MTR, RBP4, APOA4, CDKN2A, FGF2, EDNRB, ITGA2, VLA-2, CABIN1, SHBG, HMGB1, HSP90B2P, CYP3A4, GJA1, CAV1, ESR2, LTA, GDF15, BDNF, CYP2D6, NGF, SP1, TGIF1, SRC, EGF, PIK3CG, HLA-A, KCNQ1, CNR1, FBN1, CHKA, BEST1, CTNNB1, IL2, CD36, PRKAB1, TPO, ALDH7A1, CX3CR1, TH, F9, CH1, TF, HFE, IL17A, PTEN, GSTM1, DMD, GATA4, F13A1,

			TTR, FABP4, PON3, APOC1, INSR, TNFRSF1B, HTR2A, CSF3, CYP2C9, TXN, CYP11B2, PTH, CSF2, KDR, PLA2G2A, THBS1, GCG, RHOA, ALDH2, TCF7L2, NFE2L2, NOTCH1, UGT1A1, IFNA1, PPARD, SIRT1, GNHR1, PAPPA, ARR3, NPPC, AHSP, PTK2, IL13, MTOR, ITGB2, GSTT1, IL6ST, CPB2, CYP1A2, HNF4A, SLC64A, PLA2G6, TNFSF11, SLC8A1, F2RL1, AKR1A1, ALDH9A1, BGLAP, MTTP, MTRR, SULT1A3, RAGE, C4B, P2RY12, RNLS, CREB1, POMC, RAC1, LMNA, CD59, SCM5A, CYP1B1, MIF, MMP13, TIMP2, CYP19A1, CUP21A2, PTPN22, MYH14, MBL2, SELPLG, AOC3, CTSL1, PCNA, IGF2, ITGB1, CAST, CXCL12, IGHE, KCNE1, TFRC, COL1A1, COL1A2, IL2RB, PLA2G10, ANGPT2, PROCR, NOX4, HAMP, PTPN11, SLCA1, IL2RA, CCL5, IRF1, CF:AR, CA:CA, EIF4E, GSTP1, JAK2, CYP3A5, HSPG2, CCL3, MYD88, VIP, SOAT1, ADRBK1, NR4A2, MMP8, NPR2, GCH1, EPRS, PPARGC1A, F12, PECAM1, CCL4, CERPINA34, CASR, FABP2, TTF2, PROS1, CTF1, SGCB, YME1L1, CAMP, ZC3H12A, AKR1B1, MMP7, AHR, CSF1, HDAC9, CTGF, KCNMA1, UGT1A, PRKCA, COMT, S100B, EGR1, PRL, IL15, DRD4, CAMK2G, SLC22A2, CCL11, PGF, THPO, GP6, TACR1, NTS, HNF1A, SST, KCDN1, LOC646627, TBXAS1, CUP2J2, TBXA2R, ADH1C, ALOX12, AHSG, BHMT, GJA4, SLC25A4, ACLY, ALOX5AP, NUMA1, CYP27B1, CYSLTR2, SOD3, LTC4S, UCN, GHRL, APOC2, CLEC4A, KBTBD10, TNC, TYMS, SHC1, LRP1, SOCS3, ADH1B, KLK3, HSD11B1, VKORC1, SERPINB2, TNS1, RNF19A, EPOR, ITGAM, PITX2, MAPK7, FCGR3A, LEEPR, ENG, GPX1, GOT2, HRH1, NR112, CRH, HTR1A, VDAC1, HPSE, SFTPD, TAP2, RMF123, PTK2Bm NTRK2, IL6R, ACHE, GLP1R, GHR, GSR, NQO1, NR5A1, GJB2, SLC9A1, MAOA, PCSK9, FCGR2A, SERPINF1, EDN3, UCP2, TFAP2A, C4BPA, SERPINF2, TYMP, ALPP,

			CXCR2, SLC3A3, ABCG2, ADA, JAK3, HSPA1A, FASN, FGF1, F11, ATP7A, CR1, GFPA, ROCK1, MECP2, MYLK, BCHE, LIPE, ADORA1, WRN, CXCR3, CD81, SMAD7, LAMC2, MAP3K5, CHGA, IAPP, RHO, ENPP1, PTHLH, NRG1, VEGFC, ENPEP, CEBPB, NAGLU,. F2RL3, CX3CL1, BDKRB1, ADAMTS13, ELANE, ENPP2, CISH, GAST, MYOC, ATP1A2, NF1, GJB1, MEF2A, VCL, BMPR2, TUBB, CDC42, KRT18, HSF1, MYB, PRKAA2, ROCK2, TFP1, PRKG1, BMP2, CTNND1, CTH, CTSS, VAV2, NPY2R, IGFBP2, CD28, GSTA1, PPIA, APOH, S100A8, IL11, ALOX15, FBLN1, NR1H3, SCD, GIP, CHGB, PRKCB, SRD5A1,HSD11B2, CALCRL, GALNT2, ANGPTL4, KCNN4, PIK3C2A, HBEGF, CYP7A1, HLA-DRB5, BNIP3, GCKR, S100A12, PADI4, HSPA14, CXCR1, H19, KRTAP19-3, IDDM2, RAC2, YRY1, CLOCK, NGFR, DBH, CHRNA4, CACNA1C, PRKAG2, CHAT, PTGDS, NR1H2, TEK, VEGFB, MEF2C, MAPKAPK2, TNFRSF11A, HSPA9, CYSLTR1, MAT1A, OPRL1, IMPA1, CLCN2, DLD, PSMA6, PSMB8, CHI3L1, ALDH1B1, PARP2,STAR, LBP, ABCC6, RGS2, EFNB2, GJB6, APOA2, AMPD1, DYSF, FDFT1,EMD2, CCR6, GJB3, IL1RL1, ENTPD1, BBS4, CELSR2, F11R, RAPGEF3, HYAL1, ZNF259, ATOX1, ATF6, KHK, SAT1, GGH, TIMP4, SLC4A4, PDE2A, PDE3B, FADS1, FADS2, TMSB4X, TXNIP, LIMS1, RHOB, LY96, FOXO1, PNPLA2,TRH, GJC1, S:C17A5, FTO, GJD2, PRSC1, CASP12, GPBAR1, PXK, IL33, TRIB1, PBX4, NUPR1, 15-SEP, CILP2, TERC, GGT2, MTCO1, UOX, AVP
Cataract	eye		CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA,

			CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1
CDKL-5 Deficiencies or Mediated Diseases	Brain, CNS		CDKL5
Charcot-Marie-Tooth (CMT) disease (Types 1, 2, 3, 4,)	Nervous system	Muscles (dystrophy)	PMP22 (CMT1A and E), MPZ (CMT1B), LITAF (CMT1C), EGR2 (CMT1D), NEFL (CMT1F), GJB1 (CMT1X), MFN2 (CMT2A), KIF1B (CMT2A2B), RAB7A (CMT2B), TRPV4 (CMT2C), GARS (CMT2D), NEFL (CMT2E), GAPD1 (CMT2K), HSPB8 (CMT2L), DYNC1H1, CMT2O), LRSAM1 (CMT2P), IGHMBP2 (CMT2S), MORC2 (CMT2Z), GDAP1 (CMT4A), MTMR2 or SBF2/MTMR13 (CMT4B), SH3TC2 (CMT4C), NDRG1 (CMT4D), PRX (CMT4F), FIG4 (CMT4J), NT-3
Chédiak-Higashi Syndrome	Immune system	Skin, hair, eyes, neurons	LYST
Choroidermia			CHM, REP1,
Chorioretinal atrophy	eye		PRDM13, RGR, TEAD1
Chronic Granulomatous Disease	Immune system		CYBA, CYBB, NCF1, NCF2, NCF4
Chronic Mucocutaneous Candidiasis	Immune system		AIRE, CARD9, CLEC7A IL12B, IL12B1, IL1F, IL17RA, IL17RC, RORC, STAT1, STAT3, TRAF31P2
Cirrhosis	liver		KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988
Colon cancer (Familial adenomatous polyposis (FAP) and hereditary nonpolyposis colon cancer (HNPCC))	Gastrointestinal		FAP: APC HNPCC: MSH2, MLH1, PMS2, SH6, PMS1
Combined Immunodeficiency	Immune System		IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228
Cone(-rod) dystrophy	eye		AIPL1, CRX, GUA1A, GUCY2D, PITPM3, PROM1, PRPH2, RIMS1, SEMA4A, ABCA4, ADAM9, ATF6, C21ORF2, C8ORF37, CACNA2D4, CDHR1, CERKL, CNGA3, CNGB3, CNNM4, CNAT2, IFT81, KCNV2, PDE6C, PDE6H, POC1B, RAX2, RDH5, RPGRIP1, TTLL5, RetCG1, GUCY2E
Congenital Stationary Night Blindness	eye		CABP4, CACNA1F, CACNA2D4, GNAT1, CPR179, GRK1, GRM6, LRIT3, NYX, PDE6B, RDH5, RHO, RLBP1, RPE65, SAG, SLC24A1, TRPM1,
Congenital Fructose Intolerance	Metabolism		ALDOB
Cori's Disease (Glycogen Storage Disease Type III)	Various-wherever glycogen accumulates, particularly		AGL

	liver, heart, skeletal muscle		
Corneal clouding and dystrophy	eye		APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD
Cornea plana congenital			KERA, CNA2
Cri du chat Syndrome, also known as 5p syndrome and cat cry syndrome			Deletions involving only band 5p15.2 to the entire short arm of chromosome 5, e.g. CTNND2, TERT,
Cystic Fibrosis (CF)	Lungs and respiratory system	Pancreas, liver, digestive system, reproductive system, exocrine, glands,	CTFR, ABCC7, CF, MRP7, SCNN1A, those described in WO2015157070
Diabetic nephropathy	kidney		Gremlin, 12/15- lipoxygenase, TIM44,
Dent Disease (Types 1 and 2)	Kidney		Type 1: CLCN5, Type 2: ORCL
Dentatorubro-Pallidoluysian Atrophy (DRPLA) (aka Haw River and Naito-Oyanagi Disease)	CNS, brain, muscle		Atrophin-1 and Atn1
Down Syndrome	various		Chromosome 21 trisomy
Drug Addiction	Brain		Prkce; Drd2; Drd4; ABAT; GRIA2;Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1
Duane syndrome (Types 1, 2, and 3, including subgroups A, B and C). Other names for this condition include: Duane's Retraction Syndrome (or DR syndrome), Eye Retraction Syndrome, Retraction Syndrome, Congenital retraction syndrome and Stilling-Turk-Duane Syndrome	eye		CHN1, indels on chromosomes 4 and 8
Duchenne muscular dystrophy (DMD)	muscle	Cardiovascular, respiratory	DMD, BMD, dystrophin gene, intron flanking exon 51 of DMD gene, exon 51 mutations in DMD gene, see also WO2013163628 and US Pat. Pub. 20130145487
Edward's Syndrome (Trisomy 18)			Complete or partial trisomy of chromosome 18
Ehlers-Danlos Syndrome (Types I-VI)	Various depending on type: including musculoskeletal, eye, vasculature, immune, and skin		COL5A1, COL5A2, COL1A1, COL3A1, TNXB, PLOD1, COL1A2, FKBP14 and ADAMTS2
Emery-Dreifuss muscular dystrophy	muscle		LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A
Enhanced S-Cone Syndrome	eye		NR2E3, NRL
Fabry's Disease	Various – including skin,		GLA

	eyes, and gastrointestinal system, kidney, heart, brain, nervous system		
Facioscapulohumeral muscular dystrophy	muscles		FSHMD1A, FSHD1A, FRG1,
Factor H and Factor H-like 1	blood		HF1, CFH, HUS
Factor V Leiden thrombophilia and Factor V deficiency	blood		Factor V (F5)
Factor V and Factor VII deficiency	blood		MCFD2
Factor VII deficiency	blood		F7
Factor X deficiency	blood		F10
Factor XI deficiency	blood		F11
Factor XII deficiency	blood		F12, HAF
Factor XIIIA deficiency	blood		F13A1, F13A
Factor XIIIB deficiency	blood		F13B
Familial Hypercholestereolemia	Cardiovascular system		APOB, LDLR, PCSK9
Familial Mediterranean Fever (FMF) also called recurrent polyserositis or familial paroxysmal polyserositis	Various-organs/tissues with serous or synovial membranes, skin, joints	Heart, kidney, brain/CNS, reproductive organs	MEFV
Fanconi Anemia	Various – blood (anemia), immune system, cognitive, kidneys, eyes, musculoskeletal		FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCC, FANCG, RAD51, BRCA1, BRCA2, BRIP1, BACH1, FANCJ, FANCB, FANCD1, FANCD2, FANCD, FAD, FANCE, FACE, FANCF, FANCI, ERCC4, FANCL, FANCM, PALB2, RAD51C, SLX4, UBE2T, FANCB, XRCC9, PHF9, KIAA1596
Fanconi Syndrome Types I (Childhood onset) and II (Adult Onset)	kidneys		FRTS1, GATM
Fragile X syndrome and related disorders	brain		FMR1, FMR2; FXR1; FXR2; mGLUR5
Fragile XE Mental Retardation (aka Martin Bell syndrome)	Brain, nervous system		FMR1
Friedreich Ataxia (FRDA)	Brain, nervous system	heart	FXN/X25
Fuchs endothelial corneal dystrophy	Eye		TCF4; COL8A2
Galactosemia	Carbohydrate metabolism disorder	Various-where galactose accumulates – liver, brain, eyes	GALT, GALK1, and GALE
Gastrointestinal Epithelial Cancer, GI cancer			CISH
Gaucher Disease (Types 1, 2, and 3, as well as other unusual forms that may not fit into these types)	Fat metabolism disorder	Various-liver, spleen, blood, CNS, skeletal system	GBA

Griscelli syndrome			
Glaucoma	eye		MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A, those described in WO2015153780
Glomerulo sclerosis	kidney		CC chemokine ligand 2
Glycogen Storage Diseases Types I-VI -See also Cori's Disease, Pompe's Disease, McArdle's disease, Hers Disease, and Von Gierke's disease	Metabolism Diseases		SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM, see also Cori's Disease, Pompe's Disease, McArdle's disease, Hers Disease, and Von Gierke's disease
RBC Glycolytic enzyme deficiency	blood		any mutations in a gene for an enzyme in the glycolysis pathway including mutations in genes for hexokinases I and II, glucokinase, phosphoglucose isomerase, phosphofructokinase, aldolase Bm triosephosphate isomerease, glyceraldehydee-3-phosphate dehydrogenase, phosphoglycerokinase, phosphoglycerate mutase, enolase I, pyruvate kinase
Hartnup's disease	Malabsorption disease	Various- brain, gastrointestinal, skin,	SLC6A19
Hearing Loss	ear		NOX3, Hes5, BDNF,
Hemochromatosis (HH)	Iron absorption regulation disease	Various-wherever iron accumulates, liver, heart, pancreas, joints, pituitary gland	HFE and H63D
Hemophagocytic lymphohistiocytosis disorders	blood		PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3
Hemorrhagic disorders	blood		PI, ATT, F5
Hers disease (Glycogen storage disease Type VI)	liver	muscle	PYGL
Hereditary angioedema (HAE)			kalikrein B1
Hereditary Hemorrhagic Telangiectasia (Osler-Weber-Rendu Syndrome)	Skin and mucous membranes		ACVRL1, ENG and SMAD4
Hereditary Spherocytosis	blood		NK1, EPB42, SLC4A1, SPTA1, and SPTB
Hereditary Persistence of Fetal Hemoglobin	blood		HBG1, HBG2, BCL11A, promoter region of HBG 1 and/or 2 (in the CCAAT box)
Hemophilia (hemophilia A (Classic) a B (aka Christmas disease) and C)	blood		A: FVIII, F8C, HEMA B: FVIX, HEMB C: F9, F11
Hepatic adenoma	liver		TCF1, HNF1A, MODY3
Hepatic failure, early onset, and neurologic disorder	liver		SCOD1, SCO1
Hepatic lipase deficiency	liver		LIPC

Hepatoblastoma, cancer and carcinomas	liver		CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5
Hermansky-Pudlak syndrome	Skin, eyes, blood, lung, kidneys, intestine		HPS1, HPS3, HPS4, HPS5, HPS6, HPS7, DTNBP1, BLOC1, BLOC1S2, BLOC3
HIV susceptibility or infection	Immune system		IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5), those in WO2015148670A1
Holoprosencephaly (HPE) (Alobar, Semilobar, and Lobar)	brain		ACVRL1, ENG, SMAD4
Homocystinuria	Metabolic disease	Various- connective tissue, muscles, CNS, cardiovascular system	CBS, MTHFR, MTR, MTRR, and MMADHC
HPV			HPV16 and HPV18 E6/E7
HSV1, HSV2, and related keratitis	eye		HSV1 genes (immediate early and late HSV-1 genes (UL1, 1.5, 5, 6, 8, 9, 12, 15, 16, 18, 19, 22, 23, 26, 26.5, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 42, 48, 49.5, 50, 52, 54, S6, RL2, RS1, those described in WO2015153789, WO2015153791
Hunter's Syndrome (aka Mucopolysaccharidosis type II)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDS
Huntington's disease (HD) and HD-like disorders	Brain, nervous system		HD, HTT, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17, PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2, and those described in WO2013130824, WO2015089354
Hurler's Syndrome (aka mucopolysaccharidosis type I H, MPS IH)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDUA, α-L-iduronidase
Hurler-Scheie syndrome (aka mucopolysaccharidosis type I H-S, MPS I H-S)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDUA, α-L-iduronidase
hyaluronidase deficiency (aka MPS IX)	Soft and connective tissues		HYAL1
Hyper IgM syndrome	Immune system		CD40L
Hyper- tension caused renal damage	kidney		Mineral corticoid receptor
Immunodeficiencies	Immune System		CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI
Inborn errors of metabolism: including urea cycle disorders,	Metabolism diseases, liver	Various organs and cells	See also: Carbohydrate metabolism disorders (e.g. galactosemia), Amino

organic acidemias), fatty acid oxidation defects, amino acidopathies, carbohydrate disorders, mitochondrial disorders			acid Metabolism disorders (e.g. phenylketonuria), Fatty acid metabolism (e.g. MCAD deficiency), Urea Cycle disorders (e.g. Citrullinemia), Organic acidemias (e.g. Maple Syrup Urine disease), Mitochondrial disorders (e.g. MELAS), peroxisomal disorders (e.g. Zellweger syndrome)
Inflammation	Various		IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); II-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3cl1
Inflammatory Bowel Diseases (e.g. Ulcerative Colitis and Chron's Disease)	Gastrointestinal	Joints, skin	NOD2, IRGM, LRRK2, ATG5, ATG16L1, IRGM, GATM, ECM1, CDH1, LAMB1, HNF4A, GNA12, IL10, CARD9/15. CCR6, IL2RA, MST1, TNFSF15, REL, STAT3, IL23R, IL12B, FUT2
Interstitial renal fibrosis	kidney		TGF-β type II receptor
Job's Syndrome (aka Hyper IgE Syndrome)	Immune System		STAT3, DOCK8
Juvenile Retinoschisis	eye		RS1, XLRS1
Kabuki Syndrome 1			MLL4, KMT2D
Kennedy Disease (aka Spinobulbar Muscular Atrophy)	Muscles, brain, nervous system		SBMA/SMAX1/AR
Klinefelter syndrome	Various-particularly those involved in development of male characteristics		Extra X chromosome in males
Lafora Disease	Brain, CNS		EMP2A and EMP2B
Leber Congenital Amaurosis	eye		CRB1, RP12, CORD2, CRD, CRX, IMPDH1, OTX2, AIPL1, CABP4, CCT2, CEP290, CLUAP1, CRB1, CRX, DTHD1, GDF6, GUCY2D, IFT140, IQCB1, KCNJ13, LCA5, LRAT, NMNAT1, PRPH2, RD3, RDH12, RPE65, RP20, RPGRIP1, SPATA7, TULP1, LCA1, LCA4, GUC2D, CORD6, LCA3,
Lesch-Nyhan Syndrome	Metabolism disease	Various – joints, cognitive, brain, nervous system	HPRT1
Leukocyte deficiencies and disorders	blood		ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4
Leukemia	Blood		TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG,

			KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN
Limb-girdle muscular dystrophy diseases	muscle		LGMD
Lowe syndrome	brain, eyes, kidneys		OCRL
Lupus glomerulo- nephritis	kidney		MAPK1
Machado- Joseph's Disease (also known as Spinocerebellar ataxia Type 3)	Brain, CNS, muscle		ATX3
Macular degeneration	eye		ABC4, CBC1, CHM1, APOE, C1QTNF5, C2, C3, CCL2, CCR2, CD36, CFB, CFH, CFHR1, CFHR3, CNGB3, CP, CRP, CST3, CTSD, CX3CR1, ELOVL4, ERCC6, FBLN5, FBLN6, FSCN2, HMCN1, HTRA1, IL6, IL8, PLEKHA1, PROM1, PRPH2, RPGR, SERPING1, TCOF1, TIMP3, TLR3
Macular Dystrophy	eye		BEST1, C1QTNF5, CTNNA1, EFEMP1, ELOVL4, FSCN2, GUCA1B, HMCN1, IMPG1, OTX2, PRDM13, PROM1, PRPH2, RP1L1, TIMP3, ABCA4, CFH, DRAM2, IMG1, MFSD8, ADMD, STGD2, STGD3, RDS, RP7, PRPH, AVMD, AOFMD, VMD2
Malattia Leventinesse	eye		EFEMP1, FBLN3
Maple Syrup Urine Disease	Metabolism disease		BCKDHA, BCKDHB, and DBT
Marfan syndrome	Connective tissue	Musculoskeletal	FBN1
Maroteaux-Lamy Syndrome (aka MPS VI)	Musculoskeletal system, nervous system	Liver, spleen	ARSB
McArdle's Disease (Glycogen Storage Disease Type V)	Glycogen storage disease	muscle	PYGM
Medullary cystic kidney disease	kidney		UMOD, HNFJ, FJHN, MCKD2, ADMCKD2
Metachromatic leukodystrophy	Lysosomal storage disease	Nervous system	ARSA
Methylmalonic acidemia (MMA)	Metabolism disease		MMAA, MMAB, MUT, MMACHC, MMADHC, LMBRD1

Morquio Syndrome (aka MPS IV A and B)	Connective tissue, skin, bone, eyes	heart	GALNS
Mucopolysaccharidosis diseases (Types I H/S, I H, II, III A B and C, I S, IVA and B, IX, VII, and VI)	Lysosomal storage disease – affects various organs/tissues		See also Hurler/Scheie syndrome, Hurler disease, Sanfillipo syndrome, Scheie syndrome, Morquio syndrome, hyaluronidase deficiency, Sly syndrome, and Maroteaux-Lamy syndrome
Muscular Atrophy	muscle		VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1
Muscular dystrophy	muscle		FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1
Myotonic dystrophy (Type 1 and Type 2)	Muscles	Eyes, heart, endocrine	CNBP (Type 2) and DMPK (Type 1)
Neoplasia			PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR gamma; WT1 (Wilms Tumor); FGF Receptor Family members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR (Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1 (4 variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor; Bax; Bcl2; caspases family (9 members: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Neurofibromatosis (NF) (NF1, formerly Recklinghausen's NF, and NF2)	brain, spinal cord, nerves, and skin		NF1, NF2
Niemann-Pick Lipidosis (Types A, B, and C)	Lysosomal Storage Disease	Various- where sphingomyelin accumulates, particularly	Types A and B: SMPD1; Type C: NPC1 or NPC2

		spleen, liver, blood, CNS	
Noonan Syndrome	Various – musculoskeletal, heart, eyes, reproductive organs, blood		PTPN11, SOS1, RAF1 and KRAS
Norrie Disease or X-linked Familial Exudative Vitreoretinopathy	eye		NDP
North Carolina Macular Dystrophy	eye		MCDR1
Osteogenesis imperfecta (OI) (Types I, II, III, IV, V, VI, VII)	bones, musculoskeletal		COL1A1, COL1A2, CRTAP, P3H
Osteopetrosis	bones		LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1
Patau's Syndrome (Trisomy 13)	Brain, heart, skeletal system		Additional copy of chromosome 13
Parkinson's disease (PD)	Brain, nervous system		SNCA (PARK1), UCHL1 (PARK 5), and LRRK2 (PARK8), (PARK3), PARK2, PARK4, PARK7 (PARK7), PINK1 (PARK6); x-Synuclein, DJ-1, Parkin, NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, NCAP, PRKN, PDJ, DBH, NDUFV2
Pattern Dystrophy of the RPE	eye		RDS/peripherin
Phenylketonuria (PKU)	Metabolism disorder	Various due to build-up of phenylalanine, phenyl ketones in tissues and CNS	PAH, PKU1, QDPR, DHPR, PTS
Polycystic kidney and hepatic disease	Kidney, liver		FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63
Pompe's Disease	Glycogen storage disease	Various – heart, liver, spleen	GAA
Porphyria (actually refers to a group of different diseases all having a specific heme production process abnormality)	Various-wherever heme precursors accumulate		ALAD, ALAS2, CPOX, FECH, HMBS, PPOX, UROD, or UROS
posterior polymorphous corneal dystrophy	eyes		TCF4; COL8A2
Primary Hyperoxaluria (e.g. type 1)	Various – eyes, heart, kidneys, skeletal system		LDHA (lactate dehydrogenase A) and hydroxyacid oxidase 1 (HAO1)
Primary Open Angle Glaucoma (POAG)	eyes		MYOC
Primary sclerosing cholangitis	Liver, gallbladder		TCF4; COL8A2
Progeria (also called Hutchinson-Gilford progeria syndrome)	All		LMNA
Prader-Willi Syndrome	Musculoskeletal system, brain, reproductive		Deletion of region of short arm of chromosome 15, including UBE3A

	and endocrine system		
Prostate Cancer	prostate		HOXB13, MSMB, GPRC6A, TP53
Pyruvate Dehydrogenase Deficiency	Brain, nervous system		PDHA1
Kidney/Renal carcinoma	kidney		RLIP76, VEGF
Rett Syndrome	Brain		MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1
Retinitis pigmentosa (RP)	eye		ADIPOR1, ABCA4, AGBL5, ARHGEF18, ARL2BP, ARL3, ARL6, BEST1, BBS1, BBS2, C2ORF71, C8ORF37, CA4, CERKL, CLRN1, CNGA1, CMGB1, CRB1, CRX, CYP4V2, DHDDS, DHX38, EMC1, EYS, FAM161A, FSCN2, GPR125, GUCA1B, HK1, HPRPF3, HGSNAT, IDH3B, IMPDH1, IMPG2, IFT140, IFT172, KLHL7, KIAA1549, KIZ, LRAT, MAK, MERTK, MVK, NEK2, NUROD1, NR2E3, NRL, OFD1, PDE6A, PDE6B, PDE6G, POMGNT1, PRCD, PROM1, PRPF3, PRPF4, PRPF6, PRPF8, PRPF31, PRPH2, RPB3, RDH12, REEP6, RP39, RGR, RHO, RLBP1, ROM1, RP1, RP1L1, RPY, RP2, RP9, RPE65, RPGR, SAMD11, SAG, SEMA4A, SLC7A14, SNRNP200, SPP2, SPATA7, TRNT1, TOPORS, TTC8, TULP1, USH2A, ZFN408, ZNF513, see also 20120204282
Scheie syndrome (also known as mucopolysaccharidosis type I S(MPS I-S))	Various- liver, spleen, eye, joint, heart, brain, skeletal		IDUA, α-L-iduronidase
Schizophrenia	Brain		Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin); Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2 Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a; GSK3b; 5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA; DTNBP1; Dao (Dao1); TCF4; COL8A2
Secretase Related Disorders	Various		APH-1 (alpha and beta); PSEN1; NCSTN; PEN-2; Nos1, Parp1, Nat1, Nat2, CTSB, APP, APH1B, PSEN2, PSENEN, BACE1, ITM2B, CTSD, NOTCH1, TNF, INS, DYT10, ADAM17, APOE, ACE, STN, TP53, IL6, NGFR, IL1B, ACHE, CTNNB1, IGF1, IFNG, NRG1, CASP3, MAPK1, CDH1, APBB1, HMGCR,

			CREB1, PTGS2, HES1, CAT, TGFB1, ENO2, ERBB4, TRAPPC10, MAOB, NGF, MMP12, JAG1, CD40LG, PPARG, FGF2, LRP1, NOTCH4, MAPK8, PREP, NOTCH3, PRNP, CTSG, EGF, REN, CD44, SELP, GHR, ADCYAP1, INSR, GFAP, MMP3, MAPK10, SP1, MYC, CTSE, PPARA, JUN, TIMP1, IL5, IL1A, MMP9, HTR4, HSPG2, KRAS, CYCS, SMG1, IL1R1, PROK1, MAPK3, NTRK1, IL13, MME, TKT, CXCR2, CHRM1, ATXN1,PAWR, NOTCJ2, M6PR, CYP46A1, CSNK1D, MAPK14, PRG2, PRKCA, L1 CAM, CD40, NR1I2, JAG2, CTNND1, CMA1, SORT1, DLK1, THEM4, JUP, CD46, CCL11, CAV3, RNASE3, HSPA8, CASP9, CYP3A4, CCR3, TFAP2A, SCP2, CDK4, JOF1A, TCF7L2, B3GALTL, MDM2, RELA, CASP7, IDE, FANP4, CASK, ADCYAP1R1, ATF4, PDGFA, C21ORF33, SCG5, RMF123, NKFB1, ERBB2, CAV1, MMP7, TGFA, RXRA, STX1A, PSMC4, P2RY2, TNFRSF21, DLG1, NUMBL, SPN, PLSCR1, UBQLN2, UBQLN1, PCSK7, SPON1, SILV, QPCT, HESS, GCC1
Selective IgA Deficiency	Immune system		Type 1: MSH5; Type 2: TNFRSF13B
Severe Combined Immunodeficiency (SCID) and SCID-X1, and ADA-SCID	Immune system		JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4, those identified in US Pat. App. Pub. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937;
Sickle cell disease	blood		HBB, BCL11A, BCL11Ae, cis-regulatory elements of the B-globin locus, HBG 1/2 promoter, HBG distal CCAAT box region between -92 and -130 of the HBG Transcription Start Site, those described in WO2015148863, WO 2013/126794, US Pat. Pub. 20110182867
Sly Syndrome (aka MPS VII)			GUSB
Spinocerebellar Ataxias (SCA types 1, 2, 3, 6, 7, 8, 12 and 17)			ATXN1, ATXN2, ATX3
Sorsby Fundus Dystrophy	eye		TIMP3
Stargardt disease	eye		ABCR, ELOVL4, ABCA4, PROM1
Tay-Sachs Disease	Lysosomal Storage disease	Various – CNS, brain, eye	HEX-A
Thalassemia (Alpha, Beta, Delta)	blood		HBA1, HBA2 (Alpha), HBB (Beta), HBB and HBD (delta), LCRB, BCL11A, BCL11Ae, cis-regulatory

			elements of the B-globin locus, HBG 1/2 promoter, those described in WO2015148860, US Pat. Pub. 20110182867, 2015/148860
Thymic Aplasia (DiGeorge Syndrome;22q11.2 deletion syndrome)	Immune system, thymus		deletion of 30 to 40 genes in the middle of chromosome 22 at a location known as 22q11.2, including TBX1, DGCR8
Transthyretin amyloidosis (ATTR)	liver		TTR (transthyretin)
trimethylaminuria	Metabolism disease		FMO3
Trinucleotide Repeat Disorders (generally)	Various		HTT; SBMA/SMAX1/AR; FXN/X25 ATX3; ATXN1; ATXN2; DMPK; Atrophin-1 and Atn1 (DRPLA Dx); CBP (Creb-BP - global instability); VLDLR; Atxn7; Atxn10; FEN1, TNRC6A, PABPN1, JPH3, MED15, ATXN1, ATXN3, TBP, CACNA1A, ATXN80S, PPP2R2B, ATXN7, TNRC6B, TNRC6C, CELF3, MAB21L1, MSH2, TMEM185A, SIX5, CNPY3, RAXE, GNB2, RPL14, ATXN8, ISR, TTR, EP400, GIGYF2, OGG1, STC1, CNDP1, C10ORF2, MAML3, DKC1, PAXIP1, CASK, MAPT, SP1, POLG, AFF2, THBS1, TP53, ESR1, CGGBP1, ABT1, KLK3, PRNP, JUN, KCNN3, BAX, FRAXA, KBTBD10, MBNL1, RAD51, NCOA3, ERDA1, TSC1, COMP, GGLC, RRAD, MSH3, DRD2, CD44, CTCF, CCND1, CLSPN, MEF2A, PTPRU, GAPDH, TRIM22, WT1, AHR, GPX1, TPMT, NDP, ARX, TYR, EGR1, UNG, NUMBL, FABP2, EN2, CRYGC, SRP14, CRYGB, PDCD1,HOXA1, ATXN2L, PMS2, GLA, CBL, FTH1, IL12RB2, OTX2, HOXA5, POLG2, DLX2, AHRR, MANF, RMEM158, see also 20110016540
Turner's Syndrome (XO)	Various – reproductive organs, and sex characteristics, vasculature		Monosomy X
Tuberous Sclerosis	CNS, heart, kidneys		TSC1, TSC2
Usher syndrome (Types I, II, and III)	Ears, eyes		ABHD12, CDH23, CIB2, CLRN1, DFNB31, GPR98, HARS, MYO7A, PCDH15, USH1C, USH1G, USH2A, USH11A, those described in WO2015134812A1

Velocardiofacial syndrome (aka 22q11.2 deletion syndrome, DiGeorge syndrome, conotruncal anomaly face syndrome (CTAF), autosomal dominant Opitz G/BB syndrome or Cayler cardiofacial syndrome)	Various – skeletal, heart, kidney, immune system, brain		Many genes are deleted, COM, TBX1, and other are associated with symptoms
Von Gierke's Disease (Glycogen Storage Disease type I)	Glycogen Storage disease	Various – liver, kidney	G6PC and SLC37A4
Von Hippel-Lindau Syndrome	Various – cell growth regulation disorder	CNS, Kidney, Eye, visceral organs	VHL
Von Willebrand Disease (Types I, II and III)	blood		VWF
Wilson Disease	Various - Copper Storage Disease	Liver, brains, eyes, other tissues where copper builds up	ATP7B
Wiskott-Aldrich Syndrome	Immune System		WAS
Xeroderma Pigmentosum	Skin	Nervous system	POLH
XXX Syndrome	Endocrine, brain		X chromosome trisomy

[0759] In one embodiment, the compositions, systems, or components thereof can be used treat or prevent a disease in a subject by modifying one or more genes associated with one or more cellular functions, such as any one or more of those in Table 7. In one embodiment, the disease is a genetic disease or disorder. In some of embodiments, the composition, system, or component thereof can modify one or more genes or polynucleotides associated with one or more genetic diseases such as any set forth in Table 7.

Table 7. Exemplary Genes controlling Cellular Functions

CELLULAR FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1; AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2; ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3; PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A; CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1; PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK; HSP90AA1; RPS6KB1

ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1; PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A; PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN; EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1; CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor Signaling	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1; MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I; PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2; MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1; MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13; RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
Axonal Guidance Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12; IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2; ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA

Ephrin Receptor Signaling Actin Cytoskeleton Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4; AKT3; SGK
	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1; PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6; ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8; F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1; MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK
Huntington's Disease Signaling	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST; GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2; HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF; RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2; CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2; BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK; CASP3; BIRC3; PARP1

B Cell Receptor Signaling	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9; EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN; GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Leukocyte Extravasation Signaling	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA; RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB; MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2; CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8; CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRAF; GSK3B; AKT3
Acute Phase Response Signaling	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14; PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1; TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN; AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11; MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA; CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR; RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1; NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2; GSK3B; AKT3; FOXO1; CASP3; RPS6KB1

p53 Signaling Aryl Hydrocarbon Receptor Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A; BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1; PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9; CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A; HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1; SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B; BAX; AKT3
	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1; NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1; SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1; MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1; SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF; CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1; CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1; HSP90AA1
Xenobiotic Metabolism Signaling	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1; CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1; NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1; HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1; GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA; FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1; GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2; PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1; CRKL; BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN; RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2; ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8; IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A; NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1; TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1; ADIPOQ

NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1
Neuregulin Signaling Wnt & Beta catenin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HSP90AA1; RPS6KB1
	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A; WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1; PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2
Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1

NRF2-mediated Oxidative Stress Response	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1; NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL; NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Hepatic Fibrosis/Hepatic Stellate Cell Activation	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF; SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX; IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB; NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3; NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2; PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA; MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD; MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3; VAV3; PRKCA
G-Protein Coupled Receptor Signaling	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB; PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3; PRKCA
Inositol Phosphate Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC; PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2; PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC; JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF; AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN; VEGFA; AKT3; FOXO1; PRKCA

Natural Killer Cell Signaling	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11; KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6; PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: G1/S Checkpoint Regulation	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC; ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11; HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1; E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1; GSK3B; RBL1; HDAC6
T Cell Receptor Signaling	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS; NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10; JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD; FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB; CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3; BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; HGF
GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A; STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1
Amyotrophic Lateral Sclerosis Signaling	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1; PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1; APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3; STAT1
Nicotinate and Nicotinamide Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1; PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK

Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1; MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS; STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2; JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term Depression	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS; PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor Signaling	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination Pathway	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4; CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1; JUN; IL1R1; IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1; NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LRP5; CEBPB; FOXO1; PRKCA
TGF-beta Signaling	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll-like Receptor Signaling	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK; NFKB1; TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF; STAT1
Neurotrophin/TRK Signaling	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A; RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; CDC42; JUN; ATF4

FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8; APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A; TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Synaptic Long Term Potentiation	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1; PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS; PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1; ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1; CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11; HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4; HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1; STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Hypoxia Signaling in the Cardiovascular System	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT; HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM; VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 Mediated Inhibition of RXR Function	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1; MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2; TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA; NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1; SREBF1; IL1R1; CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2; CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1; PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1; PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1; FRAP1; AKT3; RPS6KB1
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC; CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A; PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in the Cardiovascular System	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3; CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1; VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4; PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C; NT5E; POLD1; NME1
cAMP-mediated Signaling	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3; SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Mitochondrial Dysfunction Notch Signaling	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9; PARK7; PSEN1; PARK2; APP; CASP3 HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4

Endoplasmic Reticulum Stress Pathway Pyrimidine Metabolism	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4; EIF2AK3; CASP3 NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B; NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7; PARK2; CASP3
Cardiac & Beta Adrenergic Signaling	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC; PPP2R5C
Glycolysis/Gluconeogenesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Interferon Signaling	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sonic Hedgehog Signaling	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glycerophospholipid Metabolism	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Phospholipid Degradation	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision Repair Pathway	ERCC5; ERCC4; XPA; XPC; ERCC1
Starch and Sucrose Metabolism	UCHL1; HK2; GCK; GPI; HK1
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Circadian Rhythm Signaling	CSNK1E; CREB1; ATF4; NR1D1
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor Signaling	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDH1A1; PKM2; LDHA
Arginine and Proline Metabolism	ALDH1A1; NOS3; NOS2A
Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose Metabolism	HK2; GCK; HK1
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Coumarine and Lignin Biosynthesis	PRDX6; PRDX1; TYR
Antigen Presentation Pathway	CALR; B2M
Biosynthesis of Steroids	NQO1; DHCR7
Butanoate Metabolism	ALDH1A1; NLGN1
Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid Metabolism	PRDX6; CHKA

Histidine Metabolism	PRMT5; ALDH1A1
Inositol Metabolism	ERO1L; APEX1
Metabolism of Xenobiotics by Cytochrome p450	GSTP1; CYP1B1
Methane Metabolism	PRDX6; PRDX1
Phenylalanine Metabolism	PRDX6; PRDX1
Propanoate Metabolism	ALDH1A1; LDHA
Selenoamino Acid Metabolism	PRMT5; AHCY
Sphingolipid Metabolism	SPHK1; SPHK2
Aminophosphonate Metabolism	PRMT5
Androgen and Estrogen Metabolism	PRMT5
Ascorbate and Aldarate Metabolism	ALDH1A1
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor Signaling	GNB2L1
NRF2-mediated Oxidative Stress Response	PRDX1
Pentose Phosphate Pathway	GPI
Pentose and Glucuronate Interconversions	UCHL1
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5, TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and Isoleucine Degradation	ALDH1A1
Glycine, Serine and Threonine Metabolism	CHKA
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental Neurology	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin; Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8; Reelin; Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

[0760] In an aspect, the invention provides a method of individualized or personalized treatment of a genetic disease in a subject in need of such treatment comprising: (a) introducing one or more mutations *ex vivo* in a tissue, organ or a cell line, or *in vivo* in a transgenic non-human mammal, comprising delivering to cell(s) of the tissue, organ, cell or mammal a composition comprising the particle delivery system or the delivery system or the virus particle of any one of the above embodiment or the cell of any one of the above embodiment, wherein the specific mutations or precise sequence substitutions are or have been correlated to the genetic disease; (b) testing treatment(s) for the genetic disease on the cells to which the vector has been delivered that have the specific mutations or precise sequence substitutions correlated to the genetic disease; and (c) treating the subject based on results from the testing of treatment(s) of step (b).

Infectious Diseases

[0761] In one embodiment, the composition, system(s), or component(s) thereof can be used to diagnose, prognose, treat, and/or prevent an infectious disease caused by a microorganism, such as bacteria, virus, fungi, parasites, or combinations thereof.

[0762] In one embodiment, the system(s) or component(s) thereof can be capable of targeting specific microorganism within a mixed population. Exemplary methods of such techniques are described in e.g. Gomaa AA, Klumpe HE, Luo ML, Selle K, Barrangou R, Beisel CL. 2014. Programmable removal of bacterial strains by use of genome-targeting composition, systems, mBio 5:e00928-13; Citorik RJ, Mimee M, Lu TK. 2014. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat Biotechnol 32:1141–1145, the teachings of which can be adapted for use with the compositions, systems, and components thereof described herein.

[0763] In one embodiment, the composition, system(s), and/or components thereof can be capable of targeting pathogenic and/or drug-resistant microorganisms, such as bacteria, virus, parasites, and fungi. In one embodiment, the composition, system(s), and/or components thereof can be capable of targeting and modifying one or more polynucleotides in a pathogenic microorganism such that the microorganism is less virulent, killed, inhibited, or is otherwise rendered incapable of causing disease and/or infecting and/or replicating in a host cell.

[0764] In one embodiment, the pathogenic bacteria that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, those of the genus Actinomyces (e.g. *A. israelii*), Bacillus (e.g. *B. anthracis, B.*

cereus), Bactereoides (e.g. *B. fragilis*), Bartonella (*B. henselae, B. quintana*), Bordetella (*B. pertussis*), Borrelia (e.g. *B. burgdorferi, B. garinii, B. afzelii, and B. recurreentis*), Brucella (e.g. *B. abortus, B. canis, B. melitensis, and B. suis*), Campylobacter (e.g. *C. jejuni),* Chlamydia (e.g. *C. pneumoniae and C. trachomatis*), Chlamydophila (e.g. *C. psittaci*), Clostridium (e.g. *C. botulinum, C. difficile, C. perfringens. C. tetani*), Corynebacterium (e.g. *C. diptheriae*), Enterococcus (e.g. *E. Faecalis, E. faecium*), Ehrlichia (*E. canis and E. chaffensis*) Escherichia (e.g. *E. coli),* Francisella (e.g. *F. tularensis*), Haemophilus (e.g. *H. influenzae*), Helicobacter (*H. pylori*), Klebsiella (E.g. *K. pneumoniae*), Legionella (e.g. *L. pneumophila*), Leptospira (e.g. *L. interrogans, L. santarosai, L. weilii, L. noguchii*), Listereia (e.g. *L. monocytogeenes*), Mycobacterium (e.g. *M. leprae, M. tuberculosis, M. ulcerans*), Mycoplasma (*M. pneumoniae*), Neisseria (*N. gonorrhoeae and N. menigitidis*), Nocardia (e.g. *N. asteeroides*), Pseudomonas (*P. aeruginosa*), Rickettsia (*R. rickettsia*), Salmonella (*S. typhi and S. typhimurium*), Shigella (*S. sonnei and S. dysenteriae*), Staphylococcus (*S. aureus, S. epidermidis*, and *S. saprophyticus*), Streeptococcus (*S. agalactiaee, S. pneumoniae, S. pyogenes*), Treponema (*T. pallidum*), Ureeaplasma (e.g. *U. urealyticum*), Vibrio (e.g. *V. cholerae*), Yersinia (e.g. *Y. pestis*, *Y. enteerocolitica*, and *Y. pseudotuberculosis*).

[0765] In one embodiment, the pathogenic virus that can be targeted and/or modified by the composition, system(s), and/or component(s) thereof described herein include, but are not limited to, a double-stranded DNA virus, a partly double-stranded DNA virus, a single-stranded DNA virus, a positive single-stranded RNA virus, a negative single-stranded RNA virus, or a double stranded RNA virus. In one embodiment, the pathogenic virus can be from the family Adenoviridae (e.g. Adenovirus), Herpeesviridae (e.g. Herpes simplex, type 1, Herpes simplex, type 2, Varicella-zoster virus, Epstein–Barr virus, Human cytomegalovirus, Human herpesvirus, type 8), Papillomaviridae (e.g. Human papillomavirus), Polyomaviridae (e.g. BK virus, JC virus), Poxviridae (e.g. smallpox), Hepadnaviridae (e.g. Hepatitis B), Parvoviridae (e.g. Parvovirus B19), Astroviridae (e.g. Human astrovirus), Caliciviridae (e.g. Norwalk virus), Picornaviridae (e.g. coxsackievirus, hepatitis A virus, poliovirus, rhinovirus), Coronaviridae (e.g. Severe acute respiratory syndrome-related coronavirus, strains: Severe acute respiratory syndrome virus, Severe acute respiratory syndrome coronavirus 2 (COVID-19)), Flaviviridae (e.g. Hepatitis C virus, yellow fever virus, dengue virus, West Nile virus,TBE virus), Togaviridae (e.g. Rubella virus), Hepeviridae (e.g. Hepatitis E virus), Retroviridae (Human immunodeficiency virus (HIV)), Orthomyxoviridae (e.g. Influenza virus), Arenaviridae (e.g.

Lassa virus), Bunyaviridae (e.g. Crimean-Congo hemorrhagic fever virus, Hantaan virus), Filoviridae (e.g. Ebola virus and Marburg virus), Paramyxoviridae (e.g. Measles virus, Mumps virus, Parainfluenza virus, Respiratory syncytial virus), Rhabdoviridae (Rabies virus), Hepatits D virus, Reoviridae (e.g. Rotavirus, Orbivirus, Coltivirus, Banna virus).

[0766] In one embodiment, the pathogenic fungi that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, those of the genus Candida (e.g. C. albicans), Aspergillus (e.g. A. fumigatus, A. flavus, A. clavatus), Cryptococcus (e.g. C. neoformans, C. gattii), Histoplasma (H. capsulatum), Pneumocystis (e.g. P. jiroveecii), Stachybotrys (e.g. S. chartarum).

[0767] In one embodiment, the pathogenic parasites that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof described herein include, but are not limited to, protozoa, helminths, and ectoparasites. In one embodiment, the pathogenic protozoa that can be targeted and/or modified by the composition, system(s), and/or component(s) thereof described herein include, but are not limited to, those from the groups Sarcodina (e.g. ameba such as Entamoeba), Mastigophora (e.g. flagellates such as Giardia and Leishmania), Cilophora (e.g. ciliates such as Balantidum), and sporozoa (e.g. plasmodium and cryptosporidium). In one embodiment, the pathogenic helminths that can be targeted and/or modified by the composition, system(s), and/or component(s) thereof described herein include, but are not limited to, flatworms (platyhelminths), thorny-headed worms (acanthoceephalins), and roundworms (nematodes). In one embodiment, the pathogenic ectoparasites that can be targeted and/or modified by the composition, system(s), and/or component(s) thereof described herein include, but are not limited to, ticks, fleas, lice, and mites.

[0768] In one embodiment, the pathogenic parasite that can be targeted and/or modified by the composition, system(s), and/or component(s) thereof described herein include, but are not limited to, Acanthamoeba spp., Balamuthia mandrillaris, Babesiosis spp. (e.g. *Babesia B. divergens, B. bigemina, B. equi, B. microfti, B. duncani*), Balantidiasis spp. (e.g. *Balantidium coli*), Blastocystis spp., Cryptosporidium spp., Cyclosporiasis spp. (e.g. *Cyclospora cayetanensis*), Dientamoebiasis spp. (e.g. *Dientamoeba fragilis*), Amoebiasis spp. (e.g. *Entamoeba histolytica*), Giardiasis spp. (e.g. *Giardia lamblia*), Isosporiasis spp. (e.g. *Isospora belli*), Leishmania spp., Naegleria spp. (e.g. *Naegleria fowleri*), Plasmodium spp. (e.g. *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale curtisi, Plasmodium ovale wallikeri, Plasmodium malariae, Plasmodium knowlesi*), Rhinosporidiosis spp. (e.g.

Rhinosporidium seeberi), Sarcocystosis spp. (e.g. S*arcocystis bovihominis, Sarcocystis suihominis*), Toxoplasma spp. (e.g. *Toxoplasma gondii*), Trichomonas spp. (e.g. *Trichomonas vaginalis*), Trypanosoma spp. (e.g. *Trypanosoma brucei*), Trypanosoma spp. (e.g. *Trypanosoma cruzi*), Tapeworm (e.g. C*estoda, Taenia multiceps, Taenia saginata, Taenia solium*), Diphyllobothrium latum spp., Echinococcus spp. (e.g. *Echinococcus granulosus, Echinococcus multilocularis, E. vogeli, E. oligarthrus*), Hymenolepis spp. (e.g. *Hymenolepis nana, Hymenolepis diminuta*), Bertiella spp. (e.g. *Bertiella mucronata, Bertiella studeri*), Spirometra (e.g. *Spirometra erinaceieuropaei*), Clonorchis spp. (e.g. *Clonorchis sinensis; Clonorchis viverrini*), Dicrocoelium spp. (e.g. *Dicrocoelium dendriticum*), Fasciola spp. (e.g. *Fasciola hepatica, Fasciola gigantica*), Fasciolopsis spp. (e.g. *Fasciolopsis buski*), Metagonimus spp. (e.g. *Metagonimus yokogawai*), Metorchis spp. (e.g. *Metorchis conjunctus*), Opisthorchis spp. (e.g. *Opisthorchis viverrini, Opisthorchis felineus*), Clonorchis spp. (e.g. *Clonorchis sinensis*), Paragonimus spp. (e.g. *Paragonimus westermani; Paragonimus africanus; Paragonimus caliensis; Paragonimus kellicotti; Paragonimus skrjabini; Paragonimus uterobilateralis*), Schistosoma sp., Schistosoma spp. (e.g. *Schistosoma mansoni, Schistosoma haematobium, Schistosoma japonicum, Schistosoma mekongi, and Schistosoma intercalatum*), Echinostoma spp. (e.g. *E. echinatum*), Trichobilharzia spp. (e.g. *Trichobilharzia regent*), Ancylostoma spp. (e.g. *Ancylostoma duodenale*), Necator spp. (e.g. *Necator americanus)*, Angiostrongylus spp., Anisakis spp., Ascaris spp. (e.g. *Ascaris lumbricoides*), Baylisascaris spp. (e.g. *Baylisascaris procyonis*), Brugia spp. (e.g. *Brugia malayi, Brugia timori*), Dioctophyme spp. (e.g. *Dioctophyme renale*), Dracunculus spp. (e.g. *Dracunculus medinensis*), Enterobius spp. (e.g. *Enterobius vermicularis, Enterobius gregorii*), Gnathostoma spp. (e.g. *Gnathostoma spinigerum, Gnathostoma hispidum*), Halicephalobus spp. (e.g. *Halicephalobus gingivalis*), Loa loa spp. (e.g. *Loa loa filaria*), Mansonella spp. (e.g. *Mansonella streptocerca*), Onchocerca spp. (e.g. *Onchocerca volvulus*), Strongyloides spp. (e.g. *Strongyloides stercoralis*), Thelazia spp. (e.g. *Thelazia californiensis, Thelazia callipaeda*), Toxocara spp. (e.g. *Toxocara canis, Toxocara cati, Toxascaris leonine*), Trichinella spp. (e.g. *Trichinella spiralis, Trichinella britovi, Trichinella nelsoni, Trichinella nativa*), Trichuris spp. (e.g. *Trichuris trichiura, Trichuris vulpis*), Wuchereria spp. (e.g. *Wuchereria bancrofti*), Dermatobia spp. (e.g. *Dermatobia hominis*), Tunga spp. (e.g. *Tunga penetrans*), Cochliomyia spp. (e.g. *Cochliomyia hominivorax*), Linguatula spp. (e.g. *Linguatula serrata*), Archiacanthocephala sp., Moniliformis sp. (e.g. *Moniliformis*

moniliformis), Pediculus spp. (e.g. *Pediculus humanus capitis, Pediculus humanus humanus*), Pthirus spp. (e.g. *Pthirus pubis*), Arachnida spp. (e.g. *Trombiculidae, Ixodidae, Argaside*), Siphonaptera spp (e.g. *Siphonaptera: Pulicinae*), Cimicidae spp. (e.g. *Cimex lectularius* and *Cimex hemipterus*), Diptera spp., Demodex spp. (e.g. *Demodex folliculorum/brevis/canis*), Sarcoptes spp. (e.g. *Sarcoptes scabiei*), Dermanyssus spp. (e.g. *Dermanyssus gallinae*), Ornithonyssus spp. (e.g. *Ornithonyssus sylviarum, Ornithonyssus bursa, Ornithonyssus bacoti*), Laelaps spp. (e.g. *Laelaps echidnina*), Liponyssoides spp. (e.g. *Liponyssoides sanguineus*).

[0769] In one embodiment the gene targets can be any of those as set forth in Table 1 of Strich and Chertow. 2019. J. Clin. Microbio. 57:4 e01307-18, which is incorporated herein as if expressed in its entirety herein.

[0770] In one embodiment, the method can include delivering a composition, system, and/or component thereof to a pathogenic organism described herein, allowing the composition, system, and/or component thereof to specifically bind and modify one or more targets in the pathogenic organism, whereby the modification kills, inhibits, reduces the pathogenicity of the pathogenic organism, or otherwise renders the pathogenic organism non-pathogenic. In one embodiment, delivery of the composition, system, occurs *in vivo* (i.e., in the subject being treated). In one embodiment occurs by an intermediary, such as microorganism or phage that is non-pathogenic to the subject but is capable of transferring polynucleotides and/or infecting the pathogenic microorganism. In one embodiment, the intermediary microorganism can be an engineered bacteria, virus, or phage that contains the composition, system(s), and/or component(s) thereof and/or vectors and/or vector systems. The method can include administering an intermediary microorganism containing the composition, system(s), and/or component(s) thereof and/or vectors and/or vector systems to the subject to be treated. The intermediary microorganism can then produce the compositions and/or component thereof or transfer a composition, system, polynucleotide to the pathogenic organism. In embodiments, where the compositions and/or component thereof, vector, or vector system is transferred to the pathogenic microorganism, the composition, system, or component thereof is then produced in the pathogenic microorganism and modifies the pathogenic microorganism such that it is less virulent, killed, inhibited, or is otherwise rendered incapable of causing disease and/or infecting and/or replicating in a host or cell thereof.

[0771] In one embodiment, where the pathogenic microorganism inserts its genetic material into the host cell's genome (e.g., a virus), the composition, system, can be designed such that it modifies the host cell's genome such that the viral DNA or cDNA cannot be replicated by the host cell's machinery into a functional virus. In one embodiment, where the pathogenic microorganism inserts its genetic material into the host cell's genome (e.g., a virus), the composition, system, can be designed such that it modifies the host cell's genome such that the viral DNA or cDNA is deleted from the host cell's genome.

[0772] It will be appreciated that inhibiting or killing the pathogenic microorganism, the disease and/or condition that its infection causes in the subject can be treated or prevented. Thus, also provided herein are methods of treating and/or preventing one or more diseases or symptoms thereof caused by any one or more pathogenic microorganisms, such as any of those described herein.

Mitochondrial Diseases

[0773] Some of the most challenging mitochondrial disorders arise from mutations in mitochondrial DNA (mtDNA), a high copy number genome that is maternally inherited. In one embodiment, mtDNA mutations can be modified using a composition, system, described herein. In one embodiment, the mitochondrial disease that can be diagnosed, prognosed, treated, and/or prevented can be MELAS (mitochondrial myopathy encephalopathy, and lactic acidosis and stroke-like episodes), CPEO/PEO (chronic progressive external ophthalmoplegia syndrome/progressive external ophthalmoplegia), KSS (Kearns-Sayre syndrome), MIDD (maternally inherited diabetes and deafness), MERRF (myoclonic epilepsy associated with ragged red fibers), NIDDM (noninsulin-dependent diabetes mellitus), LHON (Leber hereditary optic neuropathy), LS (Leigh Syndrome) an aminoglycoside induced hearing disorder, NARP (neuropathy, ataxia, and pigmentary retinopathy), Extrapyramidal disorder with akinesia-rigidity, psychosis and SNHL, Nonsyndromic hearing loss a cardiomyopathy, an encephalomyopathy, Pearson's syndrome, or a combination thereof.

[0774] In one embodiment, the mtDNA of a subject can be modified *in vivo* or *ex vivo*. In one embodiment, where the mtDNA is modified *ex vivo*, after modification the cells containing the modified mitochondria can be administered back to the subject. In one embodiment, the composition, system, or component thereof can be capable of correcting an mtDNA mutation, or a combination thereof.

[0775] In one embodiment, at least one of the one or more mtDNA mutations is selected from the group consisting of: A3243G, C3256T, T3271C, G1019A, A1304T, A15533G, C1494T, C4467A, T1658C, G12315A, A3421G, A8344G, T8356C, G8363A, A13042T, T3200C, G3242A, A3252G, T3264C, G3316A, T3394C, T14577C, A4833G, G3460A, G9804A, G11778A, G14459A, A14484G, G15257A, T8993C, T8993G, G10197A, G13513A, T1095C, C1494T, A1555G, G1541A, C1634T, A3260G, A4269G, T7587C, A8296G, A8348G, G8363A, T9957C, T9997C, G12192A, C12297T, A14484G, G15059A, duplication of CCCCCTCCCC-tandem repeats at positions 305-314 and/or 956-965, deletion at positions from 8,469-13,447, 4,308-14,874, and/or 4,398-14,822, 961ins/delC, the mitochondrial common deletion (e.g. mtDNA 4,977 bp deletion), and combinations thereof.

[0776] In one embodiment, the mitochondrial mutation can be any mutation as set forth in or as identified by use of one or more bioinformatic tools available at Mitomap available at mitomap.org. Such tools include, but are not limited to, "Variant Search, aka Market Finder", Find Sequences for Any Haplogroup, aka "Sequence Finder", "Variant Info", "POLG Pathogenicity Prediction Server", "MITOMASTER", "Allele Search", "Sequence and Variant Downloads", "Data Downloads". MitoMap contains reports of mutations in mtDNA that can be associated with disease and maintains a database of reported mitochondrial DNA Base Substitution Diseases: rRNA/tRNA mutations.

[0777] In one embodiment, the method includes delivering a composition, system, and/or a component thereof to a cell, and more specifically one or more mitochondria in a cell, allowing the composition, system, and/or component thereof to modify one or more target polynucleotides in the cell, and more specifically one or more mitochondria in the cell. The target polynucleotides can correspond to a mutation in the mtDNA, such as any one or more of those described herein. In one embodiment, the modification can alter a function of the mitochondria such that the mitochondria functions normally or at least is/are less dysfunctional as compared to an unmodified mitochondria. Modification can occur *in vivo* or *ex vivo*. Where modification is performed *ex vivo*, cells containing modified mitochondria can be administered to a subject in need thereof in an autologous or allogenic manner.

Microbiome Modification

[0778] Microbiomes play important roles in health and disease. For example, the gut microbiome can play a role in health by controlling digestion, preventing growth of pathogenic microorganisms and have been suggested to influence mood and emotion. Imbalanced

microbiomes can promote disease and are suggested to contribute to weight gain, unregulated blood sugar, high cholesterol, cancer, and other disorders. A healthy microbiome has a series of joint characteristics that can be distinguished from non-healthy individuals, thus detection and identification of the disease-associated microbiome can be used to diagnose and detect disease in an individual. The compositions, systems, and components thereof can be used to screen the microbiome cell population and be used to identify a disease associated microbiome. Cell screening methods utilizing compositions, systems, and components thereof are described elsewhere herein and can be applied to screening a microbiome, such as a gut, skin, vagina, and/or oral microbiome, of a subject.

[0779] In one embodiment, the microbe population of a microbiome in a subject can be modified using a composition, system, and/or component thereof described herein. In one embodiment, the composition, system, and/or component thereof can be used to identify and select one or more cell types in the microbiome and remove them from the microbiome population. Exemplary methods of selecting cells using a composition, system, and/or component thereof are described elsewhere herein. In this way the make-up or microorganism profile of the microbiome can be altered. In one embodiment, the alteration causes a change from a diseased microbiome composition to a healthy microbiome composition. In this way the ratio of one type or species of microorganism to another can be modified, such as going from a diseased ratio to a healthy ratio. In one embodiment, the cells selected are pathogenic microorganisms.

[0780] In one embodiment, the compositions and systems described herein can be used to modify a polynucleotide in a microorganism of a microbiome in a subject. In one embodiment, the microorganism is a pathogenic microorganism. In one embodiment, the microorganism is a commensal and non-pathogenic microorganism. Methods of modifying polynucleotides in a cell in the subject are described elsewhere herein and can be applied to these embodiments.

Models of Diseases and Conditions

[0781] In an aspect, the invention provides a method of modeling a disease associated with a genomic locus in a eukaryotic organism or a non-human organism comprising manipulation of a target sequence within a coding, non-coding or regulatory element of said genomic locus comprising delivering a non- naturally occurring or engineered composition comprising a viral vector system comprising one or more viral vectors operably encoding a composition for expression thereof, wherein the composition comprises particle delivery system or the delivery

system or the virus particle of any one of the above embodiments or the cell of any one of the above embodiment.

[0782] In one aspect, the invention provides a method of generating a model eukaryotic cell that can include one or more a mutated disease genes and/or infectious microorganisms. In one embodiment, a disease gene is any gene associated an increase in the risk of having or developing a disease. In one embodiment, the method includes (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors comprise a composition, system, and/or component thereof and/or a vector or vector system that is capable of driving expression of a composition, system, and/or component thereof including, but not limited to: a guide sequence, one or more chimeric IscB systems, and combinations thereof and (b) allowing a composition, system, or complex to bind to one or more target polynucleotides, e.g., to effect cleavage, nicking, or other modification of the target polynucleotide within said disease gene, wherein the composition, system, or complex is composed of one or more chimeric IscB systems complexed with (1) one or more ωRNAs or guide sequences that is/are hybridized to the target sequence(s) within the target polynucleotide(s), and optionally (2) the ωRNA scaffold sequence(s), thereby generating a model eukaryotic cell comprising one or more mutated disease gene(s). Thus, in one embodiment the composition and system, contains nucleic acid molecules for and drives expression of one or more of: a chimeric IscB system, a guide sequence and/or a Homologous Recombination template and/or a stabilizing ligand if the chimeric IscB system has a destabilization domain. In one embodiment, said cleavage comprises cleaving one or two strands at the location of the target sequence by the chimeric IscB systems. In one embodiment, nicking comprises nicking one or two strands at the location of the target sequence by the chimeric IscB systems. In one embodiment, said cleavage or nicking results in modified transcription of a target polynucleotide. In one embodiment, modification results in decreased transcription of the target polynucleotide. In one embodiment, the method further comprises repairing said cleaved or nicked target polynucleotide by homologous recombination with a recombination template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In one embodiment, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

[0783] The disease modeled can be any disease with a genetic or epigenetic component. In one embodiment, the disease modeled can be any as discussed elsewhere herein.

***In situ* Disease Detection**

[0784] The compositions, systems, and/or components thereof can be used for diagnostic methods of detection such as in CASFISH (see e.g., Deng et al. 2015. PNAS USA 112(38): 11870-11875), CRISPR-Live FISH (see e.g., Wang et al. 2020. Science; 365(6459):1301-1305), sm-FISH (Lee and Jefcoate. 2017. Front. Endocrinol. doi.org/10.3389/fendo.2017.00289), sequential FISH CRISPRainbow (Ma et al. Nat Biotechnol, 34 (2016), pp. 528-530), CRISPR-Sirius (Nat Methods, 15 (2018), pp. 928-931), Casilio (Cheng et al. Cell Res, 26 (2016), pp. 254-257), Halo-Tag based genomic loci visualization techniques (e.g., Deng et al. 2015. PNAS USA 112(38): 11870-11875; Knight et al., Science, 350 (2015), pp. 823-826), RNA-aptamer based methods (e.g., Ma et al., J Cell Biol, 214 (2016), pp. 529-537), molecular beacon-based methods (e.g., Zhao et al. Biomaterials, 100 (2016), pp. 172-183; Wu et al. Nucleic Acids Res (2018)), Quantum Dot-based systems (e.g., Ma et al. Anal Chem, 89 (2017), pp. 12896-12901), multiplexed methods (e.g., Ma et al., Proc Natl Acad Sci U S A, 112 (2015), pp. 3002-3007; Fu et al. Nat Commun, 7 (2016), p. 11707; Ma et al. Nat Biotechnol, 34 (2016), pp. 528-530; Shao et al. Nucleic Acids Res, 44 (2016), Article e86); Wang et al. Sci Rep, 6 (2016), p. 26857), ç, and other in situ CRISPR-hybridization based methods (e.g., Chen et al. Cell, 155 (2013), pp. 1479-1491; Gu et al. Science, 359 (2018), pp. 1050-1055; Tanebaum et al. Cell, 159 (2014), pp. 635-646; Ye et al. Protein Cell, 8 (2017), pp. 853-855; Chen et al. Nat Commun, 9 (2018), p. 5065; Shao et al. ACS Synth Biol (2017); Fu et al. Nat Commun, 7 (2016), p. 11707; Shao et al. Nucleic Acids Res, 44 (2016), Article e86; Wang et al., Sci Rep, 6 (2016), p. 26857), all of which are incorporated by reference herein as if expressed in their entirety and whose teachings can be adapted to the compositions, systems, and components thereof described herein in view of the description herein.

[0785] In one embodiment, the composition, system, or component thereof can be used in a detection method, such as an in-situ detection method described herein. In one embodiment, the composition, system, or component thereof can include a catalytically inactivate chimeric IscB system described herein and use this system in detection methods such as fluorescence in situ hybridization (FISH) or any other described herein. In one embodiment, the inactivated chimeric IscB system, which lacks the ability to produce DNA double-strand breaks may be fused with a marker, such as fluorescent protein, such as the enhanced green fluorescent protein (eEGFP) and co-expressed with small guide RNAs to target pericentric, centric and telomeric

repeats *in vivo*. The dead chimeric IscB systems or system thereof can be used to visualize both repetitive sequences and individual genes in the human genome. Such new applications of labelled dead chimeric IscB systems and compositions, systems, thereof can be important in imaging cells and studying the functional nuclear architecture, especially in cases with a small nucleus volume or complex 3-D structures.

Cell Selection

[0786] In one embodiment, the compositions, systems, and/or components thereof described herein can be used in a method to screen and/or select cells. In one embodiment, composition, system-based screening/selection method can be used to identify diseased cells in a cell population. In one embodiment, selection of the cells results in a modification in the cells such that the selected cells die. In this way, diseased cells can be identified, and removed from the healthy cell population. In one embodiment, the diseased cells can be a cancer cell, pre-cancerous cell, a virus or other pathogenic organism infected cells, or otherwise abnormal cell. In one embodiment, the modification can impart another detectable change in the cells to be selected (e.g., a functional change and/or genomic barcode) that facilitates selection of the desired cells. In one embodiment a negative selection scheme can be used to obtain a desired cell population. In these embodiments, the cells to be selected against are modified, thus can be removed from the cell population based on their death or identification or sorting based the detectable change imparted on the cells. Thus, in these embodiments, the remaining cells after selection are the desired cell population.

[0787] In one embodiment, a method of selecting one or more cell(s) containing a polynucleotide modification can include: introducing one or more composition, system,(s) and/or components thereof, and/or vectors or vector systems into the cell(s), wherein the composition, system(s), and/or components thereof, and/or vectors or vector systems contains and/or is capable of expressing one or more of: a chimeric IscB system, an ωRNA sequence, and an recombination template; wherein, for example that which is being expressed is within and expressed *in vivo* by the composition, system, vector or vector system and/or the recombination template comprises the one or more mutations that abolish chimeric IscB systems cleavage; allowing homologous recombination of the recombination template with the target polynucleotide in the cell(s) to be selected; allowing a composition, system, or complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the AAV- complex comprises the chimeric IscB systems complexed with (1) the

ωRNA or guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the ωRNA scaffold, wherein binding of the complex to the target polynucleotide induces cell death or imparts some other detectable change to the cell, thereby allowing one or more cell(s) in which one or more mutations have been introduced to be selected. In one embodiment, the cell to be selected may be a eukaryotic cell. In one embodiment, the cell to be selected may be a prokaryotic cell. Selection of specific cells via the methods herein can be performed without requiring a selection marker or a two-step process that may include a counter-selection system.

Therapeutic Agent Development

[0788] The compositions, systems, and components thereof described herein can be used to develop chimeric IscB based therapeutics. Thus, described herein are methods for developing a biologically active agent that modulates a cell function and/or signaling event associated with a disease and/or disease gene. In one embodiment, the method comprises (a) contacting a test compound with a diseased cell and/or a cell containing a disease gene cell; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event or other cell functionality associated with said disease or disease gene, thereby developing said biologically active agent that modulates said cell signaling event or other functionality associated with said disease gene. In one embodiment, the diseased cell is a model cell described elsewhere herein. In one embodiment, the diseased cell is a diseased cell isolated from a subject in need of treatment. In one embodiment, the test compound is a small molecule agent. In one embodiment, test compound is a small molecule agent. In one embodiment, the test compound is a biologic molecule agent.

[0789] In one embodiment, the method involves developing a therapeutic based on the composition, system, described herein. In one embodiment, the therapeutic comprises a chimeric IscB system and/or a ωRNA with a reprogrammable spacer capable of hybridizing to a target sequence of interest. In one embodiment, the therapeutic is a vector or vector system that can contain a) a first regulatory element operably linked to a nucleotide sequence encoding the chimeric IscB systems; and b) a second regulatory element operably linked to one or more nucleotide sequences encoding one or more nucleic acid molecules comprising a ωRNA comprising a reprogrammable spacer sequence, a conserved RNA sequence; wherein components (a) and (b) are located on same or different vectors. In one embodiment, the biologically active agent is a composition comprising a delivery system operably configured

to deliver composition, system, or components thereof, and/or or one or more polynucleotide sequences, vectors, or vector systems containing or encoding said components into a cell and capable of forming a complex with the components of the composition and system herein, and wherein said complex is operable in the cell. In one embodiment, the complex can include the chimeric IscB systems as described herein, ωRNA scaffold comprising the guide sequence (reprogrammable spacer sequence), and a conserved nucleotide sequence. In any such compositions, the delivery system can be a yeast system, a lipofection system, a microinjection system, a biolistic system, virosomes, liposomes, immunoliposomes, polycations, lipid:nucleic acid conjugates or artificial virions, or any other system as described herein. In one embodiment, the delivery is via a particle, a nanoparticle, a lipid or a cell penetrating peptide (CPP).

[0790] Also described herein are methods for developing or designing a composition, system,, optionally a composition, system, based therapy or therapeutic, comprising (a) selecting for a (therapeutic) locus of interest ωRNA target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites sub-selecting target sites, wherein a ωRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest ωRNA target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest ωRNA target sites, wherein a ωRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, and optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more ωRNA recognizing one or more of said (sub)selected target sites.

[0791] In one embodiment, the method for developing or designing a ωRNA for use in a composition, system, optionally a composition, system, based therapy or therapeutic, can include (a) selecting for a (therapeutic) locus of interest ωRNA target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites sub-selecting target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest gRNA target sites, wherein

a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more gRNA recognizing one or more of said (sub)selected target sites.

[0792] In one embodiment, thee method for developing or designing a composition, system, optionally a composition, system, based therapy or therapeutic in a population, can include (a) selecting for a (therapeutic) locus of interest reprogrammable spacer target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites sub-selecting target sites, wherein a ωRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest ω RNA reprogrammable spacer target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest ωRNA reprogrammable spacer target sites, wherein a ωRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more ωRNA recognizing one or more of said (sub)selected target sites.

[0793] In one embodiment the method for developing or designing a gRNA for use in a composition, system,, optionally a composition, system, based therapy or therapeutic in a population, can include (a) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites sub-selecting target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest gRNA target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest gRNA target sites, wherein a gRNA directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected target sites for an individual subject, optionally

designing one or more ωRNA reprogrammable spacer recognizing one or more of said (sub)selected target sites.

[0794] In one embodiment, the method for developing or designing a composition, system, such as a composition, system, based therapy or therapeutic, optionally in a population; or for developing or designing a ωRNA reprogrammable spacer for use in a composition, system, optionally a composition, system, based therapy or therapeutic, optionally in a population, can include selecting a set of target sequences for one or more loci in a target population, wherein the target sequences do not contain variants occurring above a threshold allele frequency in the target population (i.e., platinum target sequences); removing from said selected (platinum) target sequences any target sequences having high frequency off-target candidates (relative to other (platinum) targets in the set) to define a final target sequence set; preparing one or more, such as a set of compositions, systems, based on the final target sequence set, optionally wherein a number of composition prepared is based (at least in part) on the size of a target population.

[0795] In an embodiment, off-target candidates/off-targets, TAM restrictiveness, target cleavage efficiency, or effector protein specificity is identified or determined using a sequencing-based double-strand break (DSB) detection assay, such as described herein elsewhere. In an embodiment, off-target candidates/off-targets are identified or determined using a sequencing-based double-strand break (DSB) detection assay, such as described herein elsewhere. In an embodiment, off-targets, or off target candidates have at least 1, preferably 1-3, mismatches or (distal) TAM mismatches, such as 1 or more, such as 1, 2, 3, or more (distal) TAM mismatches. In an embodiment, sequencing-based DSB detection assay comprises labeling a site of a DSB with an adapter comprising a primer binding site, labeling a site of a DSB with a barcode or unique molecular identifier, or combination thereof, as described herein elsewhere.

[0796] It will be understood that the reprogrammable spacer sequence of the ωRNA is 100% complementary to the target site, i.e., does not comprise any mismatch with the target site. It will be further understood that "recognition" of an (off-)target site by a reprogrammable spacer presupposes composition, system, functionality, i.e., an (off-)target site is only recognized by a reprogrammable spacer RNA if binding of the reprogrammable spacer RNA to the (off-)target site leads to composition, system, activity (such as induction of single or double strand DNA cleavage, transcriptional modulation, etc.).

[0797] In an embodiment, the target sites having minimal sequence variation across a population are characterized by absence of sequence variation in at least 99%, preferably at least 99.9%, more preferably at least 99.99% of the population. In an embodiment, optimizing target location comprises selecting target sequences or loci having an absence of sequence variation in at least 99%, %, preferably at least 99.9%, more preferably at least 99.99% of a population. These targets are referred to herein elsewhere also as "platinum targets". In an embodiment, said population comprises at least 1000 individuals, such as at least 5000 individuals, such as at least 10000 individuals, such as at least 50000 individuals.

[0798] In an embodiment, the off-target sites are characterized by at least one mismatch between the off-target site and the ωRNA. In an embodiment, the off-target sites are characterized by at most five, preferably at most four, more preferably at most three mismatches between the off-target site and the ωRNA. In an embodiment, the off-target sites are characterized by at least one mismatch between the off-target site and the ωRNA and by at most five, preferably at most four, more preferably at most three mismatches between the off-target site and the ωRNA.

[0799] In an embodiment, said minimal number of off-target sites across said population is determined for high-frequency haplotypes in said population. In an embodiment, said minimal number of off-target sites across said population is determined for high-frequency haplotypes of the off-target site locus in said population. In an embodiment, said minimal number of off-target sites across said population is determined for high-frequency haplotypes of the target site locus in said population. In an embodiment, the high-frequency haplotypes are characterized by occurrence in at least 0.1% of the population.

[0800] In an embodiment, the number of (sub)selected target sites needed to treat a population is estimated based on based low frequency sequence variation, such as low frequency sequence variation captured in large scale sequencing datasets. In an embodiment, the number of (sub)selected target sites needed to treat a population of a given size is estimated.

[0801] In an embodiment, the method further comprises obtaining genome sequencing data of a subject to be treated; and treating the subject with a composition, system, selected from the set of compositions, systems, wherein the composition, system, selected is based (at least in part) on the genome sequencing data of the individual. In an embodiment, the ((sub)selected) target is validated by genome sequencing, preferably whole genome sequencing.

[0802] In an embodiment, target sequences or loci as described herein are (further) selected based on optimization of one or more parameters, such as TAM type (natural or modified), TAM nucleotide content, TAM length, target sequence length, TAM restrictiveness, target cleavage efficiency, and target sequence position within a gene, a locus or other genomic region. Methods of optimization are discussed in greater detail elsewhere herein.

[0803] In an embodiment, target sequences or loci as described herein are (further) selected based on optimization of one or more of target loci location, target length, target specificity, and TAM characteristics. As used herein, TAM characteristics may comprise for instance TAM sequence, TAM length, and/or TAM GC contents. In an embodiment, optimizing TAM characteristics comprises optimizing nucleotide content of a TAM. In an embodiment, optimizing nucleotide content of TAM is selecting a TAM with a motif that maximizes abundance in the one or more target loci, minimizes mutation frequency, or both. Minimizing mutation frequency can for instance be achieved by selecting TAM sequences devoid of or having low or minimal CpG.

[0804] In an embodiment, the effector protein for each composition and system, in the set of compositions, systems, is selected based on optimization of one or more parameters selected from the group consisting of; effector protein size, ability of effector protein to access regions of high chromatin accessibility, degree of uniform enzyme activity across genomic targets, epigenetic tolerance, mismatch/budge tolerance, effector protein specificity, effector protein stability or half-life, effector protein immunogenicity or toxicity. Methods of optimization are discussed in greater detail elsewhere herein.

Optimization of the Systems

[0805] The methods of the present invention can involve optimization of selected parameters or variables associated with the composition, system, and/or its functionality, as described herein further elsewhere. Optimization of the composition, system, in the methods as described herein may depend on the target(s), such as the therapeutic target or therapeutic targets, the mode or type of composition, system, modulation, such as composition, system, based therapeutic target(s) modulation, modification, or manipulation, as well as the delivery of the composition, system, components. One or more targets may be selected, depending on the genotypic and/or phenotypic outcome. For instance, one or more therapeutic targets may be selected, depending on (genetic) disease etiology or the desired therapeutic outcome. The (therapeutic) target(s) may be a single gene, locus, or other genomic site, or may be multiple

genes, loci or other genomic sites. As is known in the art, a single gene, locus, or other genomic site may be targeted more than once, such as by use of multiple ωRNAs, or ωRNA scaffold and multiple reprogrammable spacers.

[0806] The activity of the composition and/or system, such as chimeric IscB system-based therapy or therapeutics may involve target disruption, such as target mutation, such as leading to gene knockout. The activity of the composition and/or system, such as chimeric IscB system-based therapy or therapeutics may involve replacement of particular target sites, such as leading to target correction. chimeric IscB system-based therapy or therapeutics may involve removal of particular target sites, such as leading to target deletion. The activity of the composition and/or system, such as chimeric IscB system-based therapy or therapeutics may involve modulation of target site functionality, such as target site activity or accessibility, leading for instance to (transcriptional and/or epigenetic) gene or genomic region activation or gene or genomic region silencing. The skilled person will understand that modulation of target site functionality may involve chimeric IscB system mutation (such as for instance generation of a catalytically inactive chimeric IscB system) and/or functionalization (such as for instance fusion of the chimeric IscB systems with a heterologous functional domain, such as a transcriptional activator or repressor), as described herein elsewhere.

[0807] Accordingly, in an aspect, the invention relates to a method as described herein, comprising selection of one or more (therapeutic) target, selecting one or more functionality of the composition and/or system, and optimization of selected parameters or variables associated with the composition and/or its functionality. In a related aspect, the invention relates to a method as described herein, comprising (a) selecting one or more (therapeutic) target loci, (b) selecting one or more composition functionalities, (c) optionally selecting one or more modes of delivery, and preparing, developing, or designing a composition herein selected based on steps (a)-(c).

[0808] In an embodiment, the functionality of the composition and/or system comprises genomic mutation. In an embodiment, the functionality of the composition and/or system comprises single genomic mutation. In an embodiment, the functionality of the composition and/or system functionality comprises multiple genomic mutation. In an embodiment, the functionality of the composition and/or system comprises gene knockout. In an embodiment, the functionality of the composition and/or system comprises single gene knockout. In an embodiment, the functionality of the composition and/or system comprises multiple gene

knockout. In an embodiment, the functionality of the composition and/or system comprises gene correction. In an embodiment, the functionality of the composition and/or system comprises single gene correction. In an embodiment, the functionality of the composition and/or system comprises multiple gene correction. In an embodiment, the functionality of the composition and/or system comprises genomic region correction. In an embodiment, the functionality of the composition and/or system comprises single genomic region correction. In an embodiment, the functionality of the composition and/or system comprises multiple genomic region correction. In an embodiment, the functionality of the composition and/or system comprises gene deletion. In an embodiment, the functionality of the composition and/or system comprises single gene deletion. In an embodiment, the functionality of the composition and/or system comprises multiple gene deletion. In an embodiment, the functionality of the composition and/or system comprises genomic region deletion. In an embodiment, the functionality of the composition and/or system comprises single genomic region deletion. In an embodiment, the functionality of the composition and/or system comprises multiple genomic region deletion. In an embodiment, the functionality of the composition and/or system comprises modulation of gene or genomic region functionality. In an embodiment, the functionality of the composition and/or system comprises modulation of single gene or genomic region functionality. In an embodiment, the functionality of the composition and/or system comprises modulation of multiple gene or genomic region functionality. In an embodiment, the functionality of the composition and/or system comprises gene or genomic region functionality, such as gene or genomic region activity. In an embodiment, the functionality of the composition and/or system comprises single gene or genomic region functionality, such as gene or genomic region activity. In an embodiment, the functionality of the composition and/or system comprises multiple gene or genomic region functionality, such as gene or genomic region activity. In an embodiment, the functionality of the composition and/or system comprises modulation gene activity or accessibility optionally leading to transcriptional and/or epigenetic gene or genomic region activation or gene or genomic region silencing. In an embodiment, the functionality of the composition and/or system comprises modulation single gene activity or accessibility optionally leading to transcriptional and/or epigenetic gene or genomic region activation or gene or genomic region silencing. In an embodiment, the functionality of the composition and/or system comprises modulation

multiple gene activity or accessibility optionally leading to transcriptional and/or epigenetic gene or genomic region activation or gene or genomic region silencing.

[0809] Optimization of selected parameters or variables in the methods as described herein may result in optimized or improved the system, such as chimeric IscB system-based therapy or therapeutic, specificity, efficacy, and/or safety. In an embodiment, one or more of the following parameters or variables are taken into account, are selected, or are optimized in the methods of the invention as described herein: chimeric IscB system allosteric interactions, chimeric IscB system functional domains and functional domain interactions, chimeric IscB system specificity, ωRNA specificity, composition specificity, TAM restrictiveness, TAM type (natural or modified), TAM nucleotide content, TAM length, chimeric IscB system activity, ωRNA activity, chimeric IscB system/guide complex activity, target cleavage efficiency, target site selection, target sequence length, ability of effector protein to access regions of high chromatin accessibility, degree of uniform enzyme activity across genomic targets, epigenetic tolerance, mismatch/budge tolerance, chimeric IscB system stability, IscB polypeptide or chimeric IscB system mRNA stability, gRNA stability, chimeric IscB system complex stability, chimeric IscB system or mRNA immunogenicity or toxicity, gRNA immunogenicity or toxicity, chimeric IscB system immunogenicity or toxicity, chimeric IscB system or mRNA dose or titer, gRNA dose or titer, dose or titer, chimeric IscB system protein size, chimeric IscB system expression level, gRNA expression level, chimeric IscB system expression level, chimeric IscB system spatiotemporal expression, gRNA spatiotemporal expression, chimeric IscB system/ωRNA spatiotemporal expression.

[0810] By means of example, and without limitation, parameter or variable optimization may be achieved as follows. chimeric IscB system specificity may be optimized by selecting the most specific chimeric IscB system, e.g. IscB polypeptide or chimeric IscB systems. This may be achieved for instance by selecting the most specific chimeric IscB system orthologue or by specific chimeric IscB system mutations which increase specificity. ωRNA specificity may be optimized by selecting the most specific ωRNA. This can be achieved for instance by selecting ωRNA having low homology, i.e. at least one or preferably more, such as at least 2, or preferably at least 3, mismatches to off-target sites. The specificity may be optimized by increasing chimeric IscB system specificity and/or ωRNA specificity as above. TAM restrictiveness may be optimized by selecting a chimeric IscB system having to most restrictive TAM recognition. This can be achieved for instance by selecting a chimeric IscB system

ortholog having more restrictive TAM recognition or by specific chimeric IscB system mutations which increase or alter TAM restrictiveness. TAM type may be optimized for instance by selecting the appropriate chimeric IscB systems, such as the appropriate chimeric IscB systems recognizing a desired TAM type. The chimeric IscB systems or TAM type may be naturally occurring or may for instance be optimized based on chimeric IscB system mutants having an altered TAM recognition, or TAM recognition repertoire. TAM nucleotide content may for instance be optimized by selecting the appropriate chimeric IscB systems, such as the appropriate chimeric IscB system recognizing a desired TAM nucleotide content. The chimeric IscB systems or TAM type may be naturally occurring or may for instance be optimized based on chimeric IscB system mutants having an altered TAM recognition, or TAM recognition repertoire. TAM length may for instance be optimized by selecting the appropriate chimeric IscB system, such as the appropriate chimeric IscB systems recognizing a desired TAM nucleotide length. The chimeric IscB systems or TAM type may be naturally occurring or may for instance be optimized based on chimeric IscB system mutants having an altered TAM recognition, or TAM recognition repertoire.

[0811] Target length or target sequence length may be optimized, for instance, by selecting the appropriate chimeric IscB systems, such as the appropriate chimeric IscB system recognizing a desired target or target sequence nucleotide length. Alternatively, or in addition, the target (sequence) length may be optimized by providing a target having a length deviating from the target (sequence) length typically associated with the chimeric IscB system, such as the naturally occurring chimeric IscB systems. The chimeric IscB system or target (sequence) length may be naturally occurring or may for instance be optimized based on chimeric IscB system mutants having an altered target (sequence) length recognition, or target (sequence) length recognition repertoire. For instance, increasing or decreasing target (sequence) length may influence target recognition and/or off-target recognition. chimeric IscB system activity may be optimized by selecting the most active chimeric IscB systems. This may be achieved for instance by selecting the most active chimeric IscB system ortholog or by specific chimeric IscB system mutations which increase activity. The ability of the chimeric IscB systems to access regions of high chromatin accessibility, may be optimized by selecting the appropriate chimeric IscB systems or mutant thereof, and can consider the size of the chimeric IscB systems, charge, or other dimensional variables etc. The degree of uniform chimeric IscB system activity may be optimized by selecting the appropriate chimeric IscB system or mutant

thereof, and can consider chimeric IscB system specificity and/or activity, TAM specificity, target length, mismatch tolerance, epigenetic tolerance, chimeric IscB system and/or ωRNA stability and/or half-life, chimeric IscB system and/or ωRNA immunogenicity and/or toxicity, etc. ωRNA activity may be optimized by selecting the most active ωRNA. In one embodiment, this can be achieved by increasing ωRNA stability through RNA modification. compositions activity may be optimized by increasing chimeric IscB system activity and/or ωRNA activity as above.

[0812] The target site selection may be optimized by selecting the optimal position of the target site within a gene, locus or other genomic region. The target site selection may be optimized by optimizing target location comprises selecting a target sequence with a gene, locus, or other genomic region having low variability. This may be achieved for instance by selecting a target site in an early and/or conserved exon or domain (i.e., having low variability, such as polymorphisms, within a population).

[0813] In an embodiment, optimizing target (sequence) length comprises selecting a target sequence within one or more target loci between 5 and 25 nucleotides. In an embodiment, a target sequence is 20 nucleotides.

[0814] In an embodiment, optimizing target specificity comprises selecting targets loci that minimize off-target candidates.

[0815] In one embodiment, the target site may be selected by minimization of off-target effects (e.g., off-targets qualified as having 1-5, 1-4, or preferably 1-3 mismatches compared to target and/or having one or more TAM mismatches, such as distal TAM mismatches), preferably also considering variability within a population. Chimeric IscB system stability may be optimized by selecting chimeric IscB system having appropriate half-life, such as preferably a short half-life while still capable of maintaining sufficient activity. In one embodiment, this can be achieved by selecting an appropriate chimeric IscB system orthologue having a specific half-life or by specific chimeric IscB system mutations or modifications which affect half-life or stability, such as inclusion (e.g., fusion) of stabilizing or destabilizing domains or sequences. chimeric IscB system mRNA stability may be optimized by increasing or decreasing chimeric IscB system mRNA stability. In one embodiment, this can be achieved by increasing or chimeric IscB system mRNA stability through mRNA modification. ωRNA stability may be optimized by increasing or decreasing ωRNA stability. In one embodiment, this can be achieved by increasing or decreasing ωRNA stability through RNA modification. The stability

may be optimized by increasing or decreasing chimeric IscB system stability and/or gRNA stability as above. Chimeric IscB system protein or mRNA immunogenicity or toxicity may be optimized by decreasing chimeric IscB system or mRNA immunogenicity or toxicity. In one embodiment, this can be achieved by mRNA or protein modifications. Similarly, in case of DNA based expression systems, DNA immunogenicity or toxicity may be decreased. ωRNA immunogenicity or toxicity may be optimized by decreasing ωRNA immunogenicity or toxicity. In one embodiment, this can be achieved by ωRNA modifications. Similarly, in case of DNA based expression systems, DNA immunogenicity or toxicity may be decreased. The immunogenicity or toxicity may be optimized by decreasing chimeric IscB system immunogenicity or toxicity and/or ωRNA immunogenicity or toxicity as above, or by selecting the least immunogenic or toxic chimeric IscB system /ωRNA combination. Similarly, in case of DNA based expression systems, DNA immunogenicity or toxicity may be decreased. chimeric IscB system or mRNA dose or titer may be optimized by selecting dosage or titer to minimize toxicity and/or maximize specificity and/or efficacy. ωRNA dose or titer may be optimized by selecting dosage or titer to minimize toxicity and/or maximize specificity and/or efficacy. The composition dose or titer may be optimized by selecting dosage or titer to minimize toxicity and/or maximize specificity and/or efficacy. The chimeric IscB system size may be optimized by selecting minimal protein size to increase efficiency of delivery, in particular for virus mediated delivery. chimeric IscB systems, ωRNA, or complex thereof expression level may be optimized by limiting (or extending) the duration of expression and/or limiting (or increasing) expression level. This may be achieved for instance by using self-inactivating compositions, systems,, such as including a self-targeting (e.g., chimeric IscB system targeting) gRNA, by using viral vectors having limited expression duration, by using appropriate promoters for low (or high) expression levels, by combining different delivery methods for individual chimeric IscB system components, such as virus mediated delivery of chimeric IscB systems encoding nucleic acid combined with non-virus mediated delivery of ωRNA, or virus mediated delivery of ωRNA combined with non-virus mediated delivery of chimeric IscB systems or mRNA. IscB polypeptide nuclease, ωRNA, or chimeric IscB system complex spatiotemporal expression may be optimized by appropriate choice of conditional and/or inducible expression systems, including controllable chimeric IscB system activity optionally a destabilized chimeric IscB system and/or a split chimeric IscB system, and/or cell- or tissue-specific expression systems.

[0816] In an aspect, the invention relates to a method as described herein, comprising selection of one or more (therapeutic) target, selecting the functionality of the composition and/or system, selecting composition mode of delivery, selecting composition delivery vehicle or expression system, and optimization of selected parameters or variables associated with the composition and/or its functionality, optionally wherein the parameters or variables are one or more selected from chimeric IscB system specificity, ωRNA specificity, chimeric IscB system complex specificity, TAM restrictiveness, TAM type (natural or modified), TAM nucleotide content, TAM length, chimeric IscB system activity, gRNA activity, chimeric IscB system/ωRNA complex activity, target cleavage efficiency, target site selection, target sequence length, ability of effector protein to access regions of high chromatin accessibility, degree of uniform enzyme activity across genomic targets, epigenetic tolerance, mismatch/budge tolerance, chimeric IscB system stability, chimeric IscB system mRNA stability, ωRNA stability, IscB polypeptide or chimeric IscB system stability, chimeric IscB system or mRNA immunogenicity or toxicity, ωRNA immunogenicity or toxicity, chimeric IscB system/ωRNA complex immunogenicity or toxicity, chimeric IscB system or mRNA dose or titer, ωRNA dose or titer, chimeric IscB system dose or titer, chimeric IscB system size, chimeric IscB system expression level, ωRNA expression level, chimeric IscB system/ωRNA molecule complex expression level, chimeric IscB system spatiotemporal expression, ωRNA spatiotemporal expression, chimeric IscB system/ωRNA complex spatiotemporal expression.

[0817] It will be understood that the parameters or variables to be optimized as well as the nature of optimization may depend on the (therapeutic) target, the functionality of the composition and/or system, the system mode of delivery, and/or the composition delivery vehicle or expression system.

[0818] In an aspect, the invention relates to a method as described herein, comprising optimization of ωRNA specificity at the population level. Preferably, said optimization of ωRNA specificity comprises minimizing ωRNA target site sequence variation across a population and/or minimizing ωRNA off-target incidence across a population.

[0819] In one embodiment, optimization can result in selection of a chimeric IscB system that is naturally occurring or is modified. In one embodiment, optimization can result in selection of a chimeric IscB system that has nuclease, nickase, deaminase, transposase, and/or has one or more effector functionalities deactivated or eliminated. In one embodiment,

optimizing a TAM specificity can include selecting a chimeric IscB system with a modified TAM specificity. In one embodiment, optimizing can include selecting a chimeric IscB system having a minimal size. In an embodiment, optimizing effector protein stability comprises selecting an effector protein having a short half-life while maintaining sufficient activity, such as by selecting an appropriate chimeric IscB system orthologue having a specific half-life or stability. In an embodiment, optimizing immunogenicity or toxicity comprises minimizing effector protein immunogenicity or toxicity by protein modifications. In an embodiment, optimizing functional specific comprises selecting a protein effector with reduced tolerance of mismatches and/or bulges between the guide RNA and one or more target loci.

[0820] In an embodiment, optimizing efficacy comprises optimizing overall efficiency, epigenetic tolerance, or both. In an embodiment, maximizing overall efficiency comprises selecting an effector protein with uniform enzyme activity across target loci with varying chromatin complexity, selecting an effector protein with enzyme activity limited to areas of open chromatin accessibility. In an embodiment, chromatin accessibility is measured using one or more of ATAC-seq, or a DNA-proximity ligation assay. In an embodiment, optimizing epigenetic tolerance comprises optimizing methylation tolerance, epigenetic mark competition, or both. In an embodiment, optimizing methylation tolerance comprises selecting an effector protein that modify methylated DNA. In an embodiment, optimizing epigenetic tolerance comprises selecting an effector protein unable to modify silenced regions of a chromosome, selecting an effector protein able to modify silenced regions of a chromosome, or selecting target loci not enriched for epigenetic markers.

[0821] In an embodiment, selecting an optimized guide RNA comprises optimizing gRNA stability, gRNA immunogenicity, or both, or other gRNA associated parameters or variables as described herein elsewhere.

[0822] In an embodiment, optimizing gRNA stability and/or gRNA immunogenicity comprises RNA modification, or other gRNA associated parameters or variables as described herein elsewhere. In an embodiment, the modification comprises removing 1-3 nucleotides form the 3' end of a target complementarity region of the gRNA. In an embodiment, modification comprises an extended gRNA and/or trans RNA/DNA element that create stable structures in the gRNA that compete with gRNA base pairing at a target of off-target loci, or extended complimentary nucleotides between the gRNA and target sequence, or both.

[0823] In an embodiment, the mode of delivery comprises delivering gRNA and/or chimeric IscB systems, delivering gRNA and/or chimeric IscB system mRNA, or delivery gRNA and/or chimeric IscB system as a DNA based expression system. In an embodiment, the mode of delivery further comprises selecting a delivery vehicle and/or expression systems from the group consisting of liposomes, lipid particles, nanoparticles, biolistics, or viral-based expression/delivery systems. In an embodiment, expression is spatiotemporal expression is optimized by choice of conditional and/or inducible expression systems, including controllable chimeric IscB system activity optionally a destabilized chimeric IscB system and/or a split chimeric IscB system, and/or cell- or tissue-specific expression system.

[0824] The methods as described herein may further involve selection of the mode of delivery. In an embodiment, ωRNA and/or chimeric IscB systems are or are to be delivered. In an embodiment, ωRNA and/or chimeric IscB system mRNA are or are to be delivered. In an embodiment, ωRNA and/or chimeric IscB systems provided in a DNA-based expression system are or are to be delivered. In an embodiment, delivery of the individual system components comprises a combination of the above modes of delivery. In an embodiment, delivery comprises delivering ωRNA and/or chimeric IscB systems, delivering ωRNA and/or chimeric IscB system mRNA, or delivering ωRNA and/or chimeric IscB systems as a DNA based expression system.

[0825] The methods as described herein may further involve selection of the composition delivery vehicle and/or expression system. Delivery vehicles and expression systems are described herein elsewhere. By means of example, delivery vehicles of nucleic acids and/or proteins include nanoparticles, liposomes, etc. Delivery vehicles for DNA, such as DNA-based expression systems include for instance biolistics, viral based vector systems (e.g. adenoviral, AAV, lentiviral), etc. the skilled person will understand that selection of the mode of delivery, as well as delivery vehicle or expression system may depend on for instance the cell or tissues to be targeted. In an embodiment, the delivery vehicle and/or expression system for delivering the compositions, systems, or components thereof comprises liposomes, lipid particles, nanoparticles, biolistics, or viral-based expression/delivery systems.

Considerations for Therapeutic Applications

[0826] A consideration in genome editing therapy is the choice of sequence-specific nuclease, such as a variant of a chimeric IscB system. Each nuclease variant may possess its own unique set of strengths and weaknesses, many of which must be balanced in the context

of treatment to maximize therapeutic benefit. For a specific editing therapy to be efficacious, a sufficiently high level of modification must be achieved in target cell populations to reverse disease symptoms. This therapeutic modification 'threshold' is determined by the fitness of edited cells following treatment and the amount of gene product necessary to reverse symptoms. With regard to fitness, editing creates three potential outcomes for treated cells relative to their unedited counterparts: increased, neutral, or decreased fitness. In the case of increased fitness, corrected cells may be able and expand relative to their diseased counterparts to mediate therapy. In this case, where edited cells possess a selective advantage, even low numbers of edited cells can be amplified through expansion, providing a therapeutic benefit to the patient. Where the edited cells possess no change in fitness, an increase the therapeutic modification threshold can be warranted. As such, significantly greater levels of editing may be needed to treat diseases, where editing creates a neutral fitness advantage, relative to diseases where editing creates increased fitness for target cells. If editing imposes a fitness disadvantage, as would be the case for restoring function to a tumor suppressor gene in cancer cells, modified cells would be outcompeted by their diseased counterparts, causing the benefit of treatment to be low relative to editing rates. This may be overcome with supplemental therapies to increase the potency and/or fitness of the edited cells relative to the diseased counterparts.

[0827] In addition to cell fitness, the amount of gene product necessary to treat disease can also influence the minimal level of therapeutic genome editing that can treat or prevent a disease or a symptom thereof. In cases where a small change in the gene product levels can result in significant changes in clinical outcome, the minimal level of therapeutic genome editing is less relative to cases where a larger change in the gene product levels is needed to gain a clinically relevant response. In one embodiment, the minimal level of therapeutic genome editing can range from 0.1 to 1 %, 1-5%, 5-10%, 10-15%, 15-20%, 20-25%, 25-30%, 30-35%, 35-40%, 40-45%. 45-50%, or 50-55%. Thus, where a small change in gene product levels can influence clinical outcomes and diseases where there is a fitness advantage for edited cells, are ideal targets for genome editing therapy, as the therapeutic modification threshold is low enough to permit a high chance of success.

[0828] The activity of NHEJ and HDR DSB repair can vary by cell type and cell state. NHEJ is not highly regulated by the cell cycle and is efficient across cell types, allowing for high levels of gene disruption in accessible target cell populations. In contrast, HDR acts

primarily during S/G2 phase, and is therefore restricted to cells that are actively dividing, limiting treatments that require precise genome modifications to mitotic cells [Ciccia, A. & Elledge, S.J. Molecular cell 40, 179-204 (2010); Chapman, J.R., et al. Molecular cell 47, 497-510 (2012)].

[0829] The efficiency of correction via HDR may be controlled by the epigenetic state or sequence of the targeted locus, or the specific repair template configuration (single vs. double stranded, long vs. short homology arms) used [Hacein-Bey-Abina, S., et al. The New England journal of medicine 346, 1185-1193 (2002); Gaspar, H.B., et al. Lancet 364, 2181-2187 (2004); Beumer, K.J., et al. G3 (2013)]. The relative activity of NHEJ and HDR machineries in target cells may also affect gene correction efficiency, as these pathways may compete to resolve DSBs [Beumer, K.J., et al. Proceedings of the National Academy of Sciences of the United States of America 105, 19821-19826 (2008)]. HDR also imposes a delivery challenge not seen with NHEJ strategies, as it uses the concurrent delivery of nucleases and repair templates. Thus, these differences can be kept in mind when designing, optimizing, and/or selecting a chimeric IscB system based therapeutic as described in greater detail elsewhere herein.

[0830] chimeric IscB system-based polynucleotide modification application can include combinations of proteins, small RNA molecules, and/or repair templates, and can make, In one embodiment, delivery of these multiple parts substantially more challenging than, for example, traditional small molecule therapeutics. Two main strategies for delivery of compositions, systems, and components thereof have been developed: *ex vivo* and *in vivo*. In one embodiment of *ex vivo* treatments, diseased cells are removed from a subject, edited and then transplanted back into the patient. In other embodiments, cells from a healthy allogeneic donor are collected, modified using a composition or component thereof, to impart various functionalities and/or reduce immunogenicity, and administered to an allogeneic recipient in need of treatment. *Ex vivo* editing has the advantage of allowing the target cell population to be well defined and the specific dosage of therapeutic molecules delivered to cells to be specified. The latter consideration may be particularly important when off-target modifications are a concern, as titrating the amount of nuclease may decrease such mutations (Hsu et al., 2013). Another advantage of *ex vivo* approaches is the typically high editing rates that can be achieved, due to the development of efficient delivery systems for proteins and nucleic acids into cells in culture for research and gene therapy applications.

[0831] *In vivo* polynucleotide modification via compositions, systems, and/or components thereof involves direct delivery of the compositions, systems, and/or components thereof to cell types in their native tissues. *In vivo* polynucleotide modification via compositions, systems, and/or components thereof allows diseases in which the affected cell population is not amenable to *ex vivo* manipulation to be treated. Furthermore, delivering compositions, systems, and/or components thereof to cells in situ allows for the treatment of multiple tissue and cell types.

[0832] In one embodiment, such as those where viral vector systems are used to generate viral particles to deliver the composition and/or component thereof to a cell, the total cargo size of the composition and/or component thereof should be considered as vector systems can have limits on the size of a polynucleotide that can be expressed therefrom and/or packaged into cargo inside of a viral particle. In one embodiment, the tropism of a vector system, such as a viral vector system, should be considered as it can impact the cell type to which the composition or component thereof can be efficiently and/or effectively delivered.

[0833] When delivering a system or component thereof via a viral-based system, it can be important to consider the amount of viral particles that will be needed to achieve a therapeutic effect so as to account for the potential immune response that can be elicited by the viral particles when delivered to a subject or cell(s). When delivering a system or component thereof via a viral based system, it can be important to consider mechanisms of controlling the distribution and/or dosage of the system *in vivo*. Generally, to reduce the potential for off-target effects, it is optimal but not necessarily required, that the amount of the system be as close to the minimum or least effective dose. In practice this can be challenging to do.

[0834] In one embodiment, it can be important to consider the immunogenicity of the system or component thereof. In embodiments, where the immunogenicity of the system or component thereof is of concern, the immunogenicity system or component thereof can be reduced. By way of example only, the immunogenicity of the system or component thereof can be reduced using the approach set out in Tangri et al. Accordingly, directed evolution or rational design may be used to reduce the immunogenicity of the chimeric IscB system in the host species (human or other species).

Xenotransplantation

[0835] The present invention also contemplates use of the composition described herein, e.g. chimeric IscB system protein systems, to provide RNA-guided DNA nucleases adapted to

be used to provide modified tissues for transplantation. For example, RNA-guided DNA nucleases may be used to knockout, knockdown or disrupt selected genes in an animal, such as a transgenic pig (such as the human heme oxygenase-1 transgenic pig line), for example by disrupting expression of genes that encode epitopes recognized by the human immune system, i.e. xenoantigen genes. Candidate porcine genes for disruption may for example include α(l,3)-galactosyltransferase and cytidine monophosphate-N-acetylneuraminic acid hydroxylase genes (see PCT Patent Publication WO 2014/066505). In addition, genes encoding endogenous retroviruses may be disrupted, for example the genes encoding all porcine endogenous retroviruses (see Yang et al., 2015, Genome-wide inactivation of porcine endogenous retroviruses (PERVs), Science 27 November 2015: Vol. 350 no. 6264 pp. 1101-1104). In addition, RNA-guided DNA nucleases may be used to target a site for integration of additional genes in xenotransplant donor animals, such as a human CD55 gene to improve protection against hyperacute rejection.

[0836] Embodiments of the invention also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct 13, 2011 – Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep-Oct;7(5):551-8). The present effector protein systems may be harnessed to correct these defects of genomic instability.

[0837] Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

APPLICATIONS IN PLANTS AND FUNGI

[0838] The compositions, systems, and methods described herein can be used to perform gene or genome interrogation or editing or manipulation in plants and fungi. For example, the applications include investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant or fugus genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. The compositions, systems, and methods can be used with regard to plants in Site-Directed Integration (SDI) or Gene Editing (GE) or any Near Reverse Breeding (NRB) or Reverse Breeding (RB) techniques.

[0839] The compositions, systems, and methods herein may be used to confer desired traits (e.g., enhanced nutritional quality, increased resistance to diseases and resistance to biotic and abiotic stress, and increased production of commercially valuable plant products or heterologous compounds) on essentially any plants and fungi, and their cells and tissues. The compositions, systems, and methods may be used to modify endogenous genes or to modify their expression without the permanent introduction into the genome of any foreign gene.

[0840] In one embodiment, compositions, systems, and methods may be used in genome editing in plants or where RNAi or similar genome editing techniques have been used previously; see, e.g., Nekrasov, "Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR-Cas system," Plant Methods 2013, 9:39 (doi:10.1186/1746-4811-9-39); Brooks, "Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system," Plant Physiology September 2014 pp 114.247577; Shan, "Targeted genome modification of crop plants using a CRISPR-Cas system," Nature Biotechnology 31, 686-688 (2013); Feng, "Efficient genome editing in plants using a CRISPR/Cas system," Cell Research (2013) 23:1229–1232. doi:10.1038/cr.2013.114; published online 20 August 2013; Xie, "RNA-guided genome editing in plants using a CRISPR-Cas system," Mol Plant. 2013 Nov;6(6):1975-83. doi: 10.1093/mp/sst119. Epub 2013 Aug 17; Xu, "Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice," Rice 2014, 7:5 (2014), Zhou et al., "Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and Redundancy," New Phytologist (2015) (Forum) 1-4 (available online only at www.newphytologist.com); Caliando et al, "Targeted DNA degradation using a CRISPR

device stably carried in the host genome, NATURE COMMUNICATIONS 6:6989, DOI: 10.1038/ncomms7989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989; US Patent No. 6,603,061 - Agrobacterium-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, Morrell et al "Crop genomics: advances and applications," Nat Rev Genet. 2011 Dec 29;13(2):85-96, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. Aspects of utilizing the compositions, systems, and methods may be analogous to the use of the composition in plants, and mention is made of the University of Arizona website "CRISPR-PLANT" (genome.arizona.edu/crispr/) (supported by Penn State and AGI).

[0841] The compositions, systems, and methods may also be used on protoplasts. A "protoplast" refers to a plant cell that has had its protective cell wall completely or partially removed using, for example, mechanical or enzymatic means resulting in an intact biochemical competent unit of living plant that can reform their cell wall, proliferate and regenerate grow into a whole plant under proper growing conditions.

[0842] The compositions, systems, and methods may be used for screening genes (e.g., endogenous, mutations) of interest. In some examples, genes of interest include those encoding enzymes involved in the production of a component of added nutritional value or generally genes affecting agronomic traits of interest, across species, phyla, and plant kingdom. By selectively targeting e.g., genes encoding enzymes of metabolic pathways, the genes responsible for certain nutritional aspects of a plant can be identified. Similarly, by selectively targeting genes which may affect a desirable agronomic trait, the relevant genes can be identified. Accordingly, the present invention encompasses screening methods for genes encoding enzymes involved in the production of compounds with a particular nutritional value and/or agronomic traits.

[0843] It is also understood that reference herein to animal cells may also apply, *mutatis mutandis*, to plant or fungal cells unless otherwise apparent; and the enzymes herein having reduced off-target effects and systems employing such enzymes can be used in plant applications, including those mentioned herein.

[0844] In some cases, nucleic acids introduced to plants and fungi may be codon optimized for expression in the plants and fungi. Methods of codon optimization include those described

in Kwon KC, et al., Codon Optimization to Enhance Expression Yields Insights into Chloroplast Translation, Plant Physiol. 2016 Sep;172(1):62-77.

[0845] The components (e.g., chimeric IscB system) in the compositions and systems may further comprise one or more functional domains described herein. In some examples, the functional domains may be an exonuclease. Such exonuclease may increase the efficiency of the chimeric IscB system' function, e.g., mutagenesis efficiency. An example of the functional domain is Trex2, as described in Weiss T et al., www.biorxiv.org/content/10.1101/2020.04.11.037572v1, doi: doi.org/10.1101/2020.04.11.037572.

Examples of plants

[0846] The compositions, systems, and methods herein can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics. In general, the term "plant" relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants.

[0847] The compositions, systems, and methods may be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders *Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, San tales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales*; monocotyledonous plants such as those belonging to the orders *Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales*, and *Orchid ales*, or with plants belonging to *Gymnospermae*, e.g., those belonging to the orders *Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales* and *Gnetales*.

[0848] The compositions, systems, and methods herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: *Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca, Vilis, and Vigna*; and the *genera Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus*, and *Pseudotsuga*.

[0849] In one embodiment, target plants and plant cells for engineering include those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., Arabidopsis). Specifically, the plants are intended to comprise without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, eucalyptus, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, papaya, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum,

safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticale, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini.

[0850] The term plant also encompasses Algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants. The compositions, systems, and methods can be used over a broad range of "algae" or "algae cells." Examples of algae include eukaryotic phyla, including the *Rhodophyta* (red algae), *Chlorophyta* (green algae), *Phaeophyta* (brown algae), *Bacillariophyta* (diatoms), *Eustigmatophyta* and dinoflagellates as well as the prokaryotic phylum *Cyanobacteria* (blue-green algae). Examples of algae species include those of *Amphora, Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Chlorella, Chlorococcum, Cyclotella, Cylindrotheca, Dunaliella, Emiliana, Euglena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Nephrochloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova, Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraselmis, Thalassiosira, and Trichodesmium*.

Plant promoters

[0851] In order to ensure appropriate expression in a plant cell, the components of the components and systems herein may be placed under control of a plant promoter. A plant promoter is a promoter operable in plant cells. A plant promoter is capable of initiating transcription in plant cells, whether or not its origin is a plant cell. The use of different types of promoters is envisaged.

[0852] In some examples, the plant promoter is a constitutive plant promoter, which is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. In some examples, the plant promoter is a regulated promoter, which directs gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In some examples, the plant promoter is a tissue-preferred promoters, which can be utilized to target enhanced

expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed.

[0853] Exemplary plant promoters include those obtained from plants, plant viruses, and bacteria such as Agrobacterium or Rhizobium which comprise genes expressed in plant cells. Additional examples of promoters include those described in Kawamata et al., (1997) Plant Cell Physiol 38:792-803; Yamamoto et al., (1997) Plant J 12:255-65; Hire et al, (1992) Plant Mol Biol 20:207-18, Kuster et al, (1995) Plant Mol Biol 29:759-72, and Capana et al., (1994) Plant Mol Biol 25:681 -91.

[0854] In some examples, a plant promoter may be an inducible promoter, which is inducible and allows for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc.), or light inducible systems (Phytochrome, LOV domains, or cryptochrome), such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. In a particular example, of the components of a light inducible system include a chimeric IscB system, a light-responsive cytochrome heterodimer (e.g., from Arabidopsis thaliana), and a transcriptional activation/repression domain.

[0855] In some examples, the promoter may be a chemical-regulated promotor (where the application of an exogenous chemical induces gene expression) or a chemical-repressible promoter (where application of the chemical represses gene expression). Examples of chemical-inducible promoters include maize ln2-2 promoter (activated by benzene sulfonamide herbicide safeners), the maize GST promoter (activated by hydrophobic electrophilic compounds used as pre-emergent herbicides), the tobacco PR-1 a promoter (activated by salicylic acid), promoters regulated by antibiotics (such as tetracycline-inducible and tetracycline-repressible promoters).

Stable integration in the genome of plants

[0856] In one embodiment, polynucleotides encoding the components of the compositions and systems may be introduced for stable integration into the genome of a plant cell. In some cases, vectors or expression systems may be used for such integration. The design of the vector or the expression system can be adjusted depending on for when, where and under what

conditions the guide RNA and/or the chimeric IscB system gene are expressed. In some cases, the polynucleotides may be integrated into an organelle of a plant, such as a plastid, mitochondrion or a chloroplast. The elements of the expression system may be on one or more expression constructs which are either circular such as a plasmid or transformation vector, or non-circular such as linear double stranded DNA.

[0857] In one embodiment, the method of integration generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host tissue, and regenerating plant cells or plants therefrom. In some examples, the expression system for stable integration into the genome of a plant cell may contain one or more of the following elements: a promoter element that can be used to express the RNA and/or chimeric IscB system in a plant cell; a 5' untranslated region to enhance expression; an intron element to further enhance expression in certain cells, such as monocot cells; a multiple-cloning site to provide convenient restriction sites for inserting the guide RNA and/or the chimeric IscB system gene sequences and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript.

Transient expression in plants

[0858] In one embodiment, the components of the compositions and systems may be transiently expressed in the plant cell. In some examples, the compositions and systems may modify a target nucleic acid only when both the guide RNA and the chimeric IscB systems are present in a cell, such that genomic modification can further be controlled. As the expression of the chimeric IscB system is transient, plants regenerated from such plant cells typically contain no foreign DNA. In certain examples, the chimeric IscB system is stably expressed and the guide sequence is transiently expressed.

[0859] DNA and/or RNA (e.g., mRNA) may be introduced to plant cells for transient expression. In such cases, the introduced nucleic acid may be provided in sufficient quantity to modify the cell but do not persist after a contemplated period of time has passed or after one or more cell divisions.

[0860] The transient expression may be achieved using suitable vectors. Exemplary vectors that may be used for transient expression include a pEAQ vector (may be tailored for Agrobacterium-mediated transient expression) and Cabbage Leaf Curl virus (CaLCuV), and vectors described in Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93; and Yin K et al., Scientific Reports volume 5, Article number: 14926 (2015).

[0861] Combinations of the different methods described above are also envisaged.

Translocation to and/or expression in specific plant organelles

[0862] The compositions and systems herein may comprise elements for translocation to and/or expression in a specific plant organelle.

Chloroplast targeting

[0863] In one embodiment, it is envisaged that the compositions and systems are used to specifically modify chloroplast genes or to ensure expression in the chloroplast. The compositions and systems (e.g., chimeric IscB systems, ωRNAs, or their encoding polynucleotides) may be transformed, compartmentalized, and/or targeted to the chloroplast. In an example, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[0864] Examples of methods of chloroplast transformation include Particle bombardment, PEG treatment, and microinjection, and the translocation of transformation cassettes from the nuclear genome to the plastid. In some examples, targeting of chloroplasts may be achieved by incorporating in chloroplast localization sequence, and/or the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the components of the compositions and systems. Additional examples of transforming, targeting and localization of chloroplasts include those described in WO2010061186, Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology, Vol. 61: 157-180, and US 20040142476, which are incorporated by reference herein in their entireties.

Exemplary applications in plants

[0865] The compositions, systems, and methods may be used to generate genetic variation(s) in a plant (e.g., crop) of interest. One or more, e.g., a library of, ωRNAs targeting one or more locations in a genome may be provided and introduced into plant cells together with the chimeric IscB system. For example, a collection of genome-scale point mutations and gene knock-outs can be generated. In some examples, the compositions, systems, and methods may be used to generate a plant part or plant from the cells so obtained and screening the cells for a trait of interest. The target genes may include both coding and non-coding regions. In some cases, the trait is stress tolerance, and the method is a method for the generation of stress-tolerant crop varieties.

[0866] In one embodiment, the compositions, systems, and methods are used to modify endogenous genes or to modify their expression. The expression of the components may induce targeted modification of the genome, either by direct activity of the chimeric IscB systems and optionally introduction of recombination template DNA, or by modification of genes targeted. The different strategies described herein above allow chimeric IscB system -mediated targeted genome editing without requiring the introduction of the components into the plant genome.

[0867] In some cases, the modification may be performed without the permanent introduction into the genome of the plant of any foreign gene, including those encoding components of the composition herein, so as to avoid the presence of foreign DNA in the genome of the plant. This can be of interest as the regulatory requirements for non-transgenic plants are less rigorous. Components which are transiently introduced into the plant cell are typically removed upon crossing.

[0868] For example, the modification may be performed by transient expression of the components of the compositions and systems. The transient expression may be performed by delivering the components of the compositions and systems with viral vectors, delivery into protoplasts, with the aid of particulate molecules such as nanoparticles or CPPs.

Generation of plants with desired traits

[0869] The compositions, systems, and methods herein may be used to introduce desired traits to plants. The approaches include introduction of one or more foreign genes to confer a trait of interest, editing or modulating endogenous genes to confer a trait of interest.

Agronomic traits

[0870] In one embodiment, crop plants can be improved by influencing specific plant traits. Examples of the traits include improved agronomic traits such as herbicide resistance, disease resistance, abiotic stress tolerance, high yield, and superior quality, pesticide-resistance, disease resistance, insect and nematode resistance, resistance against parasitic weeds, drought tolerance, nutritional value, stress tolerance, self-pollination voidance, forage digestibility biomass, and grain yield.

[0871] In one embodiment, genes that confer resistance to pests or diseases may be introduced to plants. In cases there are endogenous genes that confer such resistance in a plants, their expression and function may be enhanced (e.g., by introducing extra copies, modifications that enhance expression and/or activity).

[0872] Examples of genes that confer resistance include plant disease resistance genes (e.g., Cf- 9, Pto, RSP2, SlDMR6-1), genes conferring resistance to a pest (e.g., those described in WO96/30517), Bacillus thuringiensis proteins, lectins, Vitamin-binding proteins (e.g., avidin), enzyme inhibitors (e.g., protease or proteinase inhibitors or amylase inhibitors), insect-specific hormones or pheromones (e.g., ecdysteroid or a juvenile hormone, variant thereof, a mimetic based thereon, or an antagonist or agonist thereof) or genes involved in the production and regulation of such hormone and pheromones, insect-specific peptides or neuropeptide, Insect-specific venom (e.g., produced by a snake, a wasp, etc., or analog thereof), Enzymes responsible for a hyperaccumulation of a monoterpene, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or another nonprotein molecule with insecticidal activity, Enzymes involved in the modification of biologically active molecule (e.g., a glycolytic enzyme, a proteolytic enzyme, a lipolytic enzyme, a nuclease, a cyclase, a transaminase, an esterase, a hydrolase, a phosphatase, a kinase, a phosphorylase, a polymerase, an elastase, a chitinase and a glucanase, whether natural or synthetic), molecules that stimulates signal transduction, Viral-invasive proteins or a complex toxin derived therefrom, Developmental-arrestive proteins produced in nature by a pathogen or a parasite, a developmental-arrestive protein produced in nature by a plant, or any combination thereof.

[0873] The compositions, systems, and methods may be used to identify, screen, introduce or remove mutations or sequences lead to genetic variability that give rise to susceptibility to certain pathogens, e.g., host specific pathogens. Such approach may generate plants that are non-host resistance, e.g., the host and pathogen are incompatible or there can be partial resistance against all races of a pathogen, typically controlled by many genes and/or also complete resistance to some races of a pathogen but not to other races.

[0874] In one embodiment, the compositions, systems, and methods may be used to modify genes involved in plant diseases. Such genes may be removed, inactivated, or otherwise regulated or modified. Examples of plant diseases include those described in [0045]-[0080] of US20140213619A1, which is incorporated by reference herein in its entirety.

[0875] In one embodiment, genes that confer resistance to herbicides may be introduced to plants. Examples of genes that confer resistance to herbicides include genes conferring resistance to herbicides that inhibit the growing point or meristem, such as an imidazolinone or a sulfonylurea, genes conferring glyphosate tolerance (e.g., resistance conferred by, e.g., mutant 5-enolpyruvylshikimate-3- phosphate synthase genes, aroA genes and glyphosate

acetyl transferase (GAT) genes, respectively), or resistance to other phosphono compounds such as by glufosinate (phosphinothricin acetyl transferase (PAT) genes from Streptomyces species, including Streptomyces hygroscopicus and Streptomyces viridichromogenes), and to pyridinoxy or phenoxy proprionic acids and cyclohexones by ACCase inhibitor-encoding genes), genes conferring resistance to herbicides that inhibit photosynthesis (such as a triazine (psbA and gs+ genes) or a benzonitrile (nitrilase gene), and glutathione S-transferase), genes encoding enzymes detoxifying the herbicide or a mutant glutamine synthase enzyme that is resistant to inhibition, genes encoding a detoxifying enzyme is an enzyme encoding a phosphinothricin acetyltransferase (such as the bar or pat protein from Streptomyces species), genes encoding hydroxyphenylpyruvatedioxygenases (HPPD) inhibitors, e.g., naturally occurring HPPD resistant enzymes, and genes encoding a mutated or chimeric HPPD enzyme.

[0876] In one embodiment, genes involved in Abiotic stress tolerance may be introduced to plants. Examples of genes include those capable of reducing the expression and/or the activity of poly(ADP-ribose) polymerase (PARP) gene, transgenes capable of reducing the expression and/or the activity of the PARG encoding genes, genes coding for a plant-functional enzyme of the nicotineamide adenine dinucleotide salvage synthesis pathway including nicotinamidase, nicotinate phosphoribosyltransferase, nicotinic acid mononucleotide adenyl transferase, nicotinamide adenine dinucleotide synthetase or nicotine amide phosphorybosyltransferase, enzymes involved in carbohydrate biosynthesis, enzymes involved in the production of polyfructose (e.g., the inulin and levan-type), the production of alpha-1,6 branched alpha-1,4-glucans, the production of alternan, the production of hyaluronan.

[0877] In one embodiment, genes that improve drought resistance may be introduced to plants. Examples of genes Ubiquitin Protein Ligase protein (UPL) protein (UPL3), DR02, DR03, ABC transporter, and DREB1A.

Nutritionally improved plants

[0878] In one embodiment, the compositions, systems, and methods may be used to produce nutritionally improved plants. In some examples, such plants may provide functional foods, e.g., a modified food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains. In certain examples, such plants may provide nutraceuticals foods, e.g., substances that may be considered a food or part of a food and provides health benefits, including the prevention and treatment of disease. The nutraceutical foods may be

useful in the prevention and/or treatment of diseases in animals and humans, e.g., cancers, diabetes, cardiovascular disease, and hypertension.

[0879] An improved plant may naturally produce one or more desired compounds and the modification may enhance the level or activity or quality of the compounds. In some cases, the improved plant may not naturally produce the compound(s), while the modification enables the plant to produce such compound(s). In some cases, the compositions, systems, and methods used to modify the endogenous synthesis of these compounds indirectly, e.g. by modifying one or more transcription factors that controls the metabolism of this compound.

[0880] Examples of nutritionally improved plants include plants comprising modified protein quality, content and/or amino acid composition, essential amino acid contents, oils and fatty acids, carbohydrates, vitamins and carotenoids, functional secondary metabolites, and minerals. In some examples, the improved plants may comprise or produce compounds with health benefits. Examples of nutritionally improved plants include those described in Newell-McGloughlin, Plant Physiology, July 2008, Vol. 147, pp. 939–953.

[0881] Examples of compounds that can be produced include carotenoids (e.g., α-Carotene or β-Carotene), lutein, lycopene, Zeaxanthin, Dietary fiber (e.g., insoluble fibers, β-Glucan, soluble fibers, fatty acids (e.g., ω-3 fatty acids, Conjugated linoleic acid, GLA,), Flavonoids (e.g., Hydroxycinnamates, flavonols, catechins and tannins), Glucosinolates, indoles, isothiocyanates (e.g., Sulforaphane), Phenolics (e.g., stilbenes, caffeic acid and ferulic acid, epicatechin), Plant stanols/sterols, Fructans, inulins, fructo-oligosaccharides, Saponins, Soybean proteins, Phytoestrogens (e.g., isoflavones, lignans), Sulfides and thiols such as diallyl sulphide, Allyl methyl trisulfide, dithiolthiones, Tannins, such as proanthocyanidins, or any combination thereof.

[0882] The compositions, systems, and methods may also be used to modify protein/starch functionality, shelf life, taste/aesthetics, fiber quality, and allergen, antinutrient, and toxin reduction traits.

[0883] Examples of genes and nucleic acids that can be modified to introduce the traits include stearyl-ACP desaturase, DNA associated with the single allele which may be responsible for maize mutants characterized by low levels of phytic acid, Tf RAP2.2 and its interacting partner SINAT2, Tf Dof1, and DOF Tf AtDof1.1 (OBP2).

Modification of polyploid plants

[0884] The compositions, systems, and methods may be used to modify polyploid plants. Polyploid plants carry duplicate copies of their genomes (e.g., as many as six, such as in wheat). In some cases, the compositions, systems, and methods may be/can be multiplexed to affect all copies of a gene, or to target dozens of genes at once. For instance, the compositions, systems, and methods may be used to simultaneously ensure a loss of function mutation in different genes responsible for suppressing defenses against a disease. The modification may be simultaneous suppression the expression of the TaMLO-Al, TaMLO-Bl and TaMLO-Dl nucleic acid sequence in a wheat plant cell and regenerating a wheat plant therefrom, in order to ensure that the wheat plant is resistant to powdery mildew (e.g., as described in WO2015109752).

Regulation of fruit-ripening

[0885] The compositions, systems, and methods may be used to regulate ripening of fruits. Ripening is a normal phase in the maturation process of fruits and vegetables. Only a few days after it starts it may render a fruit or vegetable inedible, which can bring significant losses to both farmers and consumers.

[0886] In one embodiment, the compositions, systems, and methods are used to reduce ethylene production. In some examples, the compositions, systems, and methods may be used to suppress the expression and/or activity of ACC synthase, insert a ACC deaminase gene or a functional fragment thereof, insert a SAM hydrolase gene or functional fragment thereof, suppress ACC oxidase gene expression

[0887] Alternatively or additionally, the compositions, systems, and methods may be used to modify ethylene receptors (e.g., suppressing ETR1) and/or Polygalacturonase (PG). Suppression of a gene may be achieved by introducing a mutation, an antisense sequence, and/or a truncated copy of the gene to the genome.

Increasing storage life of plants

[0888] In one embodiment, the compositions, systems, and methods are used to modify genes involved in the production of compounds which affect storage life of the plant or plant part. The modification may be in a gene that prevents the accumulation of reducing sugars in potato tubers. Upon high-temperature processing, these reducing sugars react with free amino acids, resulting in brown, bitter-tasting products and elevated levels of acrylamide, which is a potential carcinogen. In one embodiment, the methods provided herein are used to reduce or

inhibit expression of the vacuolar invertase gene (VInv), which encodes a protein that breaks down sucrose to glucose and fructose.

Reducing allergens in plants

[0889] In one embodiment, the compositions, systems, and methods are used to generate plants with a reduced level of allergens, making them safer for consumers. To this end, the compositions, systems, and methods may be used to identify and modify (e.g., suppress) one or more genes responsible for the production of plant allergens. Examples of such genes include Lol p5, as well as those in peanuts, soybeans, lentils, peas, lupin, green beans, mung beans, such as those described in Nicolaou et al., Current Opinion in Allergy and Clinical Immunology 2011;11(3):222), which is incorporated by reference herein in its entirety.

Generation of male sterile plants

[0890] The compositions, systems, and methods may be used to generate male sterile plants. Hybrid plants typically have advantageous agronomic traits compared to inbred plants. However, for self-pollinating plants, the generation of hybrids can be challenging. In different plant types (e.g., maize and rice), genes have been identified which are important for plant fertility, more particularly male fertility. Plants that are as such genetically altered can be used in hybrid breeding programs.

[0891] The compositions, systems, and methods may be used to modify genes involved male fertility, e.g., inactivating (such as by introducing mutations to) genes required for male fertility. Examples of the genes involved in male fertility include cytochrome P450-like gene (MS26) or the meganuclease gene (MS45), and those described in Wan X et al., Mol Plant. 2019 Mar 4;12(3):321-342; and Kim YJ, et al., Trends Plant Sci. 2018 Jan;23(1):53-65.

Increasing the fertility stage in plants

[0892] In one embodiment, the compositions, systems, and methods may be used to prolong the fertility stage of a plant such as of a rice. For instance, a rice fertility stage gene such as Ehd3 can be targeted in order to generate a mutation in the gene and plantlets can be selected for a prolonged regeneration plant fertility stage.

Production of early yield of products

[0893] In one embodiment, the compositions, systems, and methods may be used to produce early yield of the product. For example, flowering process may be modulated, e.g., by mutating flowering repressor gene such as SP5G. Examples of such approaches include those described in Soyk S, et al., Nat Genet. 2017 Jan;49(1):162-168.

Oil and biofuel production

[0894] The compositions, systems, and methods may be used to generate plants for oil and biofuel production. Biofuels include fuels made from plant and plant-derived resources. Biofuels may be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. Biofuels include bioethanol and biodiesel. Bioethanol can be produced by the sugar fermentation process of cellulose (starch), which may be derived from maize and sugar cane. Biodiesel can be produced from oil crops such as rapeseed, palm, and soybean. Biofuels can be used for transportation.

Generation of plants for production of vegetable oils and biofuels

[0895] The compositions, systems, and methods may be used to generate algae (e.g., diatom) and other plants (e.g., grapes) that express or overexpress high levels of oil or biofuels.

[0896] In some cases, the compositions, systems, and methods may be used to modify genes involved in the modification of the quantity of lipids and/or the quality of the lipids. Examples of such genes include those involved in the pathways of fatty acid synthesis, e.g., acetyl-CoA carboxylase, fatty acid synthase, 3-ketoacyl_acyl- carrier protein synthase III, glycerol-3-phospate deshydrogenase (G3PDH), Enoyl-acyl carrier protein reductase (Enoyl-ACP-reductase), glycerol-3-phosphate acyltransferase, lysophosphatidic acyl transferase or diacylglycerol acyltransferase, phospholipid:diacylglycerol acyltransferase, phoshatidate phosphatase, fatty acid thioesterase such as palmitoyi protein thioesterase, or malic enzyme activities.

[0897] In further embodiments, it is envisaged to generate diatoms that have increased lipid accumulation. This can be achieved by targeting genes that decrease lipid catabolization. Examples of genes include those involved in the activation of triacylglycerol and free fatty acids, β-oxidation of fatty acids, such as genes of acyl-CoA synthetase, 3-ketoacyl-CoA thiolase, acyl-CoA oxidase activity and phosphoglucomutase.

[0898] In some examples, algae may be modified for production of oil and biofuels, including fatty acids (e.g., fatty esters such as acid methyl esters (FAME) and fatty acid ethyl esters (FAEE)). Examples of methods of modifying microalgae include those described in

Stovicek et al. Metab. Eng. Comm., 2015; 2:1; US Patent No. 8,945,839; and International Patent Publication No. WO 2015/086795.

[0899] In some examples, one or more genes may be introduced (e.g., overexpressed) to the plants (e.g., algae) to produce oils and biofuels (e.g., fatty acids) from a carbon source (e.g., alcohol). Examples of the genes include genes encoding acyl-CoA synthases, ester synthases, thioesterases (e.g., tesA, 'tesA, tesB, fatB, fatB2, fatB3, fatAl, or fatA), acyl-CoA synthases (e.g., fadD, JadK, BH3103, pfl-4354, EAV15023, fadDl, fadD2, RPC_4074,fadDD35, fadDD22, faa39), ester synthases (e.g., synthase/acyl-CoA:diacylglycerl acyltransferase from *Simmondsia chinensis*, *Acinetobacter sp. ADP*, *Alcanivorax borkumensis*, *Pseudomonas aeruginosa*, *Fundibacter jadensis*, *Arabidopsis thaliana*, or *Alkaligenes eutrophus*, or variants thereof).

[0900] Additionally or alternatively, one or more genes in the plants (e.g., algae) may be inactivated (e.g., expression of the genes is decreased). For examples, one or more mutations may be introduced to the genes. Examples of such genes include genes encoding acyl-CoA dehydrogenases (e.g., fade), outer membrane protein receptors, and transcriptional regulator (e.g., repressor) of fatty acid biosynthesis (e.g., fabR), pyruvate formate lyases (e.g., pflB), lactate dehydrogenases (e.g., IdhA).

Organic acid production

[0901] In one embodiment, plants may be modified to produce organic acids such as lactic acid. The plants may produce organic acids using sugars, pentose or hexose sugars. To this end, one or more genes may be introduced (e.g., and overexpressed) in the plants. An example of such genes includes LDH gene.

[0902] In some examples, one or more genes may be inactivated (e.g., expression of the genes is decreased). For examples, one or more mutations may be introduced to the genes. The genes may include those encoding proteins involved an endogenous metabolic pathway which produces a metabolite other than the organic acid of interest and/or wherein the endogenous metabolic pathway consumes the organic acid.

[0903] Examples of genes that can be modified or introduced include those encoding pyruvate decarboxylases (pdc), fumarate reductases, alcohol dehydrogenases (adh), acetaldehyde dehydrogenases, phosphoenolpyruvate carboxylases (ppc), D-lactate dehydrogenases (d-ldh), L-lactate dehydrogenases (l-ldh), lactate 2-monooxygenases, lactate

dehydrogenase, cytochrome-dependent lactate dehydrogenases (e.g., cytochrome B2-dependent L-lactate dehydrogenases).

Enhancing plant properties for biofuel production

[0904] In one embodiment, the compositions, systems, and methods are used to alter the properties of the cell wall of plants to facilitate access by key hydrolyzing agents for a more efficient release of sugars for fermentation. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In one embodiment, lignin biosynthesis may be downregulated in the plant so as to increase fermentable carbohydrates.

[0905] In some examples, one or more lignin biosynthesis genes may be down regulated. Examples of such genes include 4-coumarate 3-hydroxylases (C3H), phenylalanine ammonia-lyases (PAL), cinnamate 4-hydroxylases (C4H), hydroxycinnamoyl transferases (HCT), caffeic acid O-methyltransferases (COMT), caffeoyl CoA 3-O-methyltransferases (CCoAOMT), ferulate 5- hydroxylases (F5H), cinnamyl alcohol dehydrogenases (CAD), cinnamoyl CoA-reductases (CCR), 4- coumarate-CoA ligases (4CL), monolignol-lignin-specific glycosyltransferases, and aldehyde dehydrogenases (ALDH), and those described in WO 2008064289.

[0906] In some examples, plant mass that produces lower level of acetic acid during fermentation may be reduced. To this end, genes involved in polysaccharide acetylation (e.g., Cas1L and those described in WO 2010096488) may be inactivated.

Other microorganisms for oils and biofuel production

[0907] In one embodiment, microorganisms other than plants may be used for production of oils and biofuels using the compositions, systems, and methods herein. Examples of the microorganisms include those of the genus of *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Synechococcus*, *Synechoystis*, *Pseudomonas*, *Aspergillus*, *Trichoderma, Neurospora, Fusarium, Humicola, Rhizomucor, Kluyveromyces, Pichia, Mucor, Myceliophtora, Penicillium, Phanerochaete, Pleurotus, Trametes, Chrysosporium, Saccharomyces, Stenotrophamonas, Schizosaccharomyces, Yarrowia, or Streptomyces.*

Plant cultures and regeneration

[0908] In one embodiment, the modified plants or plant cells may be cultured to regenerate a whole plant which possesses the transformed or modified genotype and thus the desired phenotype. Examples of regeneration techniques include those relying on manipulation of certain phytohormones in a tissue culture growth medium, relying on a biocide and/or herbicide

marker which has been introduced together with the desired nucleotide sequences, obtaining from cultured protoplasts, plant callus, explants, organs, pollens, embryos or parts thereof.

Detecting modifications in the plant genome- selectable markers

[0909] When the compositions, systems, and methods are used to modify a plant, suitable methods may be used to confirm and detect the modification made in the plant. In some examples, when a variety of modifications are made, one or more desired modifications or traits resulting from the modifications may be selected and detected. The detection and confirmation may be performed by biochemical and molecular biology techniques such as Southern analysis, PCR, Northern blot, S1 RNase protection, primer-extension or reverse transcriptase-PCR, enzymatic assays, ribozyme activity, gel electrophoresis, Western blot, immunoprecipitation, enzyme-linked immunoassays, *in situ* hybridization, enzyme staining, and immunostaining.

[0910] In some cases, one or more markers, such as selectable and detectable markers, may be introduced to the plants. Such markers may be used for selecting, monitoring, isolating cells and plants with desired modifications and traits. A selectable marker can confer positive or negative selection and is conditional or non-conditional on the presence of external substrates. Examples of such markers include genes and proteins that confer resistance to antibiotics, such as hygromycin (hpt) and kanamycin (nptII), and genes that confer resistance to herbicides, such as phosphinothricin (bar) and chlorosulfuron (als), enzyme capable of producing or processing a colored substances (e.g., the β-glucuronidase, luciferase, B or C1 genes).

Applications in fungi

[0911] The compositions, systems, and methods described herein can be used to perform efficient and cost-effective gene or genome interrogation or editing or manipulation in fungi or fungal cells, such as yeast. The approaches and applications in plants may be applied to fungi as well.

[0912] A fungal cell may be any type of eukaryotic cell within the kingdom of fungi, such as phyla of *Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia*, and *Neocallimastigomycota*. Examples of fungi or fungal cells in include yeasts, molds, and filamentous fungi.

[0913] In one embodiment, the fungal cell is a yeast cell. A yeast cell refers to any fungal cell within the phyla *Ascomycota and Basidiomycota*. Examples of yeasts include budding yeast, fission yeast, and mold, *S. cerervisiae, Kluyveromyces marxianus, Issatchenkia*

orientalis, *Candida spp.* (e.g., *Candida albicans*), *Yarrowia spp.* (e.g., *Yarrowia lipolytica*), *Pichia spp.* (e.g., *Pichia pastoris*), *Kluyveromyces spp.* (e.g., *Kluyveromyces lactis* and *Kluyveromyces marxianus*), *Neurospora spp.* (e.g., *Neurospora crassa*), *Fusarium spp.* (e.g., *Fusarium oxysporum*), and *Issatchenkia spp.* (e.g., *Issatchenkia orientalis*, *Pichia kudriavzevii* and *Candida acidothermophilum*).

[0914] In one embodiment, the fungal cell is a filamentous fungal cell, which grow in filaments, e.g., hyphae or mycelia. Examples of filamentous fungal cells include *Aspergillus spp.* (e.g., *Aspergillus niger*), *Trichoderma spp.* (e.g., *Trichoderma reesei*), *Rhizopus spp.* (e.g., *Rhizopus oryzae*), and *Mortierella spp.* (e.g., *Mortierella isabellina*).

[0915] In one embodiment, the fungal cell is of an industrial strain. Industrial strains include any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains include, without limitation, JAY270 and ATCC4124.

[0916] In one embodiment, the fungal cell is a polyploid cell whose genome is present in more than one copy. Polyploid cells include cells naturally found in a polyploid state, and cells that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may be a cell whose entire genome is polyploid, or a cell that is polyploid in a particular genomic locus of interest. In some examples, the abundance of guide RNA may more often be a rate-limiting component in genome engineering of polyploid cells than in haploid cells, and thus the methods using the composition described herein may take advantage of using certain fungal cell types.

[0917] In one embodiment, the fungal cell is a diploid cell, whose genome is present in two copies. Diploid cells include cells naturally found in a diploid state, and cells that have been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest.

[0918] In one embodiment, the fungal cell is a haploid cell, whose genome is present in one copy. Haploid cells include cells naturally found in a haploid state, or cells that have been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0919] The compositions and systems, and nucleic acid encoding thereof may be introduced to fungi cells using the delivery systems and methods herein. Examples of delivery systems include lithium acetate treatment, bombardment, electroporation, and those described in Kawai et al., 2010, Bioeng Bugs. 2010 Nov-Dec; 1(6): 395–403.

[0920] In some examples, a yeast expression vector (e.g., those with one or more regulatory elements) may be used. Examples of such vectors include a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

Biofuel and materials production by fungi

[0921] In one embodiment, the compositions, systems, and methods may be used for generating modified fungi for biofuel and material productions. For instance, the modified fungi for production of biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. Foreign genes required for biofuel production and synthesis may be introduced in to fungi In some examples, the genes may encode enzymes involved in the conversion of pyruvate to ethanol or another product of interest, degrade cellulose (e.g., cellulase), endogenous metabolic pathways which compete with the biofuel production pathway.

[0922] In some examples, the compositions, systems, and methods may be used for generating and/or selecting yeast strains with improved xylose or cellobiose utilization, isoprenoid biosynthesis, and/or lactic acid production. One or more genes involved in the metabolism and synthesis of these compounds may be modified and/or introduced to yeast

cells. Examples of the methods and genes include lactate dehydrogenase, PDC1 and PDC5, and those described in Ha, S.J., et al. (2011) Proc. Natl. Acad. Sci. USA 108(2):504-9 and Galazka, J.M., et al. (2010) Science 330(6000):84-6; Jakočiūnas T et al., Metab Eng. 2015 Mar;28:213-222; Stovicek V, et al., FEMS Yeast Res. 2017 Aug 1;17(5).

Improved plants and yeast cells

[0923] The present disclosure further provides improved plants and fungi. The improved and fungi may comprise one or more genes introduced, and/or one or more genes modified by the compositions, systems, and methods herein. The improved plants and fungi may have increased food or feed production (e.g., higher protein, carbohydrate, nutrient or vitamin levels), oil and biofuel production (e.g., methanol, ethanol), tolerance to pests, herbicides, drought, low or high temperatures, excessive water, etc.

[0924] The plants or fungi may have one or more parts that are improved, e.g., leaves, stems, roots, tubers, seeds, endosperm, ovule, and pollen. The parts may be viable, nonviable, regeneratable, and/or non- regeneratable.

[0925] The improved plants and fungi may include gametes, seeds, embryos, either zygotic or somatic, progeny and/or hybrids of improved plants and fungi. The progeny may be a clone of the produced plant or fungi, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly plants.

Further applications in plants

[0926] Further applications of the compositions, systems, and methods on plants and fungi include visualization of genetic element dynamics (e.g., as described in Chen B, et al., Cell. 2013 Dec 19;155(7):1479-91), targeted gene disruption positive-selection *in vitro* and *in vivo* (as described in Malina A et al., Genes Dev. 2013 Dec 1;27(23):2602-14), epigenetic modification such as using fusion of chimeric IscB system and histone-modifying enzymes (e.g., as described in Rusk N, Nat Methods. 2014 Jan;11(1):28), identifying transcription regulators (e.g., as described in Waldrip ZJ, Epigenetics. 2014 Sep;9(9):1207-11), anti-virus treatment for both RNA and DNA viruses (e.g., as described in Price AA, et al., Proc Natl Acad Sci U S A. 2015 May 12;112(19):6164-9; Ramanan V et al., Sci Rep. 2015 Jun 2;5:10833), alteration of genome complexity such as chromosome numbers (e.g., as described in Karimi-Ashtiyani R et al., Proc Natl Acad Sci U S A. 2015 Sep 8;112(36):11211-6; Anton T, et al., Nucleus. 2014 Mar-Apr;5(2):163-72), self-cleavage of the composition for controlled

inactivation/activation (e.g., as described Sugano SS et al., Plant Cell Physiol. 2014 Mar;55(3):475-81), multiplexed gene editing (as described in Kabadi AM et al., Nucleic Acids Res. 2014 Oct 29;42(19):e147), development of kits for multiplex genome editing (as described in Xing HL et al., BMC Plant Biol. 2014 Nov 29;14:327), starch production (as described in Hebelstrup KH et al., Front Plant Sci. 2015 Apr 23;6:247), targeting multiple genes in a family or pathway (e.g., as described in Ma X et al., Mol Plant. 2015 Aug;8(8):1274-84), regulation of non-coding genes and sequences (e.g., as described in Lowder LG, et al., Plant Physiol. 2015 Oct;169(2):971-85), editing genes in trees (e.g., as described in Belhaj K et al., Plant Methods. 2013 Oct 11;9(1):39; Harrison MM, et al., Genes Dev. 2014 Sep 1;28(17):1859-72; Zhou X et al., New Phytol. 2015 Oct;208(2):298-301), introduction of mutations for resistance to host-specific pathogens and pests.

[0927] Additional examples of modifications of plants and fungi that may be performed using the compositions, systems, and methods include analogous modifications described in International Patent Publication Nos. WO2016/099887, WO2016/025131, WO2016/073433, WO2017/066175, WO2017/100158, WO 2017/105991, WO2017/106414, WO2016/100272, WO2016/100571, WO 2016/100568, WO 2016/100562, and WO 2017/019867.

APPLICATIONS IN NON-HUMAN ANIMALS

[0928] The compositions, systems, and methods may be used to study and modify non-human animals, e.g., introducing desirable traits and disease resilience, treating diseases, facilitating breeding, etc. In one embodiment, the compositions, systems, and methods may be used to improve breeding and introducing desired traits, e.g., increasing the frequency of trait-associated alleles, introgression of alleles from other breeds/species without linkage drag, and creation of de novo favorable alleles. Genes and other genetic elements that can be targeted may be screened and identified. Examples of application and approaches include those described in Tait-Burkard C, et al., Livestock 2.0 - genome editing for fitter, healthier, and more productive farmed animals. Genome Biol. 2018 Nov 26;19(1):204; Lillico S, Agricultural applications of genome editing in farmed animals. Transgenic Res. 2019 Aug;28(Suppl 2):57-60; Houston RD, et al., Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020 Apr 16. doi: 10.1038/s41576-020-0227-y, which are incorporated herein by reference in their entireties. Applications described in other sections such as therapeutic, diagnostic, etc. can also be used on the animals herein.

[0929] The compositions, systems, and methods may be used on animals such as fish, amphibians, reptiles, mammals, and birds. The animals may be farm and agriculture animals, or pets. Examples of farm and agriculture animals include horses, goats, sheep, swine, cattle, llamas, alpacas, and birds, e.g., chickens, turkeys, ducks, and geese. The animals may be a non-human primate, e.g., baboons, capuchin monkeys, chimpanzees, lemurs, macaques, marmosets, tamarins, spider monkeys, squirrel monkeys, and vervet monkeys. Examples of pets include dogs, cats horses, wolfs, rabbits, ferrets, gerbils, hamsters, chinchillas, fancy rats, guinea pigs, canaries, parakeets, and parrots.

[0930] In one embodiment, one or more genes may be introduced (e.g., overexpressed) in the animals to obtain or enhance one or more desired traits. Growth hormones, insulin-like growth factors (IGF-1) may be introduced to increase the growth of the animals, e.g., pigs or salmon (such as described in Pursel VG et al., J Reprod Fertil Suppl. 1990;40:235-45; Waltz E, Nature. 2017;548:148). Fat-1 gene (e.g., from C elegans) may be introduced for production of larger ratio of n-3 to n-6 fatty acids may be induced, e.g. in pigs (such as described in Li M, et al., Genetics. 2018;8:1747–54). Phytase (e.g., from E coli) xylanase (e.g., from Aspergillus niger), beta-glucanase (e.g., from bacillus lichenformis) may be introduced to reduce the environmental impact through phosphorous and nitrogen release reduction, e.g. in pigs (such as described in Golovan SP, et al., Nat Biotechnol. 2001;19:741–5; Zhang X et al., elife. 2018). shRNA decoy may be introduced to induce avian influenza resilience e.g. in chicken (such as described in Lyall et al., Science. 2011;331:223–6). Lysozyme or lysostaphin may be introduced to induce mastitis resilience e.g., in goat and cow (such as described in Maga EA et al., Foodborne Pathog Dis. 2006;3:384–92; Wall RJ, et al., Nat Biotechnol. 2005;23:445–51). Histone deacetylase such as HDAC6 may be introduced to induce PRRSV resilience, e.g., in pig (such as described in Lu T., et al., PLoS One. 2017;12:e0169317). CD163 may be modified (e.g., inactivated or removed) to introduce PRRSV resilience in pigs (such as described in Prather RS et al.., Sci Rep. 2017 Oct 17;7(1):13371). Similar approaches may be used to inhibit or remove viruses and bacteria (e.g., Swine Influenza Virus (SIV) strains which include influenza C and the subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3, as well as pneumonia, meningitis and oedema) that may be transmitted from animals to humans.

[0931] In one embodiment, one or more genes may be modified or edited for disease resistance and production traits. Myostatin (e.g., GDF8) may be modified to increase muscle

growth, e.g., in cow, sheep, goat, catfish, and pig (such as described in Crispo M et al., PLoS One. 2015;10:e0136690; Wang X, et al., Anim Genet. 2018;49:43–51; Khalil K, et al., Sci Rep. 2017;7:7301; Kang J-D, et al., RSC Adv. 2017;7:12541–9). Pc POLLED may be modified to induce horlessness, e.g., in cow (such as described in Carlson DF et al., Nat Biotechnol. 2016;34:479–81). KISS1R may be modified to induce boretaint (hormone release during sexual maturity leading to undesired meat taste), e.g., in pigs. Dead end protein (dnd) may be modified to induce sterility, e.g., in salmon (such as described in Wargelius A, et al., Sci Rep. 2016;6:21284). Nano2 and DDX may be modified to induce sterility (e.g., in surrogate hosts), e.g., in pigs and chicken (such as described Park K-E, et al., Sci Rep. 2017;7:40176; Taylor L et al., Development. 2017;144:928–34). CD163 may be modified to induce PRRSV resistance, e.g., in pigs (such as described in Whitworth KM, et al., Nat Biotechnol. 2015;34:20–2). RELA may be modified to induce ASFV resilience, e.g., in pigs (such as described in Lillico SG, et al., Sci Rep. 2016;6:21645). CD18 may be modified to induce Mannheimia (Pasteurella) haemolytica resilience, e.g., in cows (such as described in Shanthalingam S, et al., roc Natl Acad Sci U S A. 2016;113:13186–90). NRAMP1 may be modified to induce tuberculosis resilience, e.g., in cows (such as described in Gao Y et al., Genome Biol. 2017;18:13). Endogenous retrovirus genes may be modified or removed for xenotransplantation such as described in Yang L, et al. Science. 2015;350:1101–4; Niu D et al., Science. 2017;357:1303–7). Negative regulators of muscle mass (e.g., Myostatin) may be modified (e.g., inactivated) to increase muscle mass, e.g., in dogs (as described in Zou Q et al., J Mol Cell Biol. 2015 Dec;7(6):580-3).

[0932] Animals such as pigs with severe combined immunodeficiency (SCID) may generated (e.g., by modifying RAG2) to provide useful models for regenerative medicine, xenotransplantation (discussed also elsewhere herein), and tumor development. Examples of methods and approaches include those described Lee K, et al., Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5; and Schomberg et al. FASEB Journal, April 2016; 30(1):Suppl 571.1.

[0933] SNPs in the animals may be modified. Examples of methods and approaches include those described Tan W. et al., Proc Natl Acad Sci U S A. 2013 Oct 8;110(41):16526-31; Mali P, et al., Science. 2013 Feb 15;339(6121):823-6.

[0934] Stem cells (e.g., induced pluripotent stem cells) may be modified and differentiated into desired progeny cells, e.g., as described in Heo YT et al., Stem Cells Dev. 2015 Feb 1;24(3):393-402.

[0935] Profile analysis (such as Igenity) may be performed on animals to screen and identify genetic variations related to economic traits. The genetic variations may be modified to introduce or improve the traits, such as carcass composition, carcass quality, maternal and reproductive traits and average daily gain.

KITS

[0936] In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In one aspect, the invention provides a kit comprising one or more of the components described herein. In one embodiment, the kit comprises the compositions herein and instructions for using the kit. In one embodiment, the kit comprises a vector system and instructions for using the kit. In one embodiment, the kit comprises a delivery system and instructions for using the kit. In one embodiment, the kit comprises a vector system and instructions for using the kit. Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube. The kits may include the ωRNA and optionally an unbound protector strand as described herein. The kits may include the ωRNA with a protector strand bound to at least partially to a reprogrammable spacer portion of the ωRNA sequence (i.e., phRNA). Thus, the kits may include the phRNA in the form of a partially double stranded nucleotide sequence as described here. In one embodiment, the kit includes instructions in one or more languages, for example in more than one language. The instructions may be specific to the applications and methods described herein.

[0937] In one embodiment, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (e.g., in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In one embodiment, the buffer is alkaline. In one embodiment, the buffer has a pH from about 7 to about 10. In one embodiment, the kit comprises one or more oligonucleotides corresponding to a ωRNA scaffold, reprogrammable sequence for insertion into a vector so as to operably link the ωRNA sequence and a regulatory element. In one embodiment, the kit comprises a homologous recombination template polynucleotide. In one

embodiment, the kit comprises one or more of the vectors and/or one or more of the polynucleotides described herein. The kit may advantageously allow to provide all elements of the systems of the invention.

[0938] The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

[0939] Further embodiments are illustrated in the following Examples which are given for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLES

Example 1. Engineered IscBs

[0940] Applicants have investigated improvement of IscB compositions by engineering of domains in the IscB polypeptide and engineering ωRNA for increased specificity and/or activity.

IscB engineered with REC domains

[0941] The REC helical bundle is involved in RNA stem binding, as shown in FIG. 1A for Cje Cas9. A small REC like insertion, shown in a large IscB (IscB-L) protein in FIG. 1B has compatible join locations that can allow for insertion of CjeCas9 REC domains. As shown in FIG. 2A the Type II Cas9 and the IscB-L have common connections to the RuvC-II domain and to the bridge helix that provide join locations for a REC domain insertion in IscB. FIG. 2B shows example junction positions at or near amino acids 153-160 on Rd8_117. Identification of secondary structures between orthologs can be used to define junctions and identify locations for potential insertions or domain replacements, which, for example can be between beta sheet alignments or loop regions. An example insertion of a Type II-D REC3 domain into IscB is shown in FIG. 3, with the interactions with ωRNA, the guide portion and target DNA depicted.

[0942] The PK-loop of ωRNA clashes with many REC domains, with designed insertion location on the IscB polypeptide having potential to clash sterically with the REC-like ωRNA region (FIG. 4A, 4B). As depicted in FIG. 4C, the highlighted REC-like ωRNA region in ribbon diagram and the clashing portion of the REC like domain can be addressed with approach to remove clashing region and rejoin with a linker. ALphafold2 models show clash free folding for an example Rd8_117 IscB polypeptide with a REC fragment from CjeCas9 inserted. (FIG. 5A, 5B). The insertion of the small region of the CjeCAs9REC in the IscB polypeptide shows a new RNA/DNA hybrid recognizing groove. (FIG. 6) A ribbon diagram of RNaseH insertion

in same region of example IscB shows fit, but that may need some rearrangement; this region of IscB can also permit other domain insertions. (FIG. 7).

IscB TAM Determining Region

[0943] An example IscB ribbon diagram shows 2 TAM determining regions, with the WED/adaptor stabilizer region and Tudor domain region highlighted. (FIG. 9). Strong polar and hydrophobic interactions suggest reduction in TAM length could present challenge in engineering. Large structural diversity exists in both the WED/adaptor stabilizer domain and the tudor core+divergent extensions. (FIG. 11).

IscB Domain Engineering Strategies

[0944] IscB polypeptide Rd8_117 has3 key contacts with TAM-proximal DNA (FIG. 12). Considering the IscB structure, some engineering strategies are identified in FIG. 13A-13F. Example IscB polypeptide approaches include full TI domain exchange with an NGG TAM/PAM (13A), RNA interaction-preserving insertions with large IscBs which can include insertion at one linker, two linkers, and/or the tudor lance region which can correspond to Rd8_117 amino acid positions 376-383, 446-448, and 462-486 (13B); exchange of Tudor Lance region of the IscB polypeptide with NGG TAM/PAMs (13C); exchange of Tudor region of the IscB polypeptide with NGG TAM/PAMs (13D); failry comprehensive insertions or exchange of the IscB polypeptide with domains of a different IscB polypeptide (13E replacement of Rd8_117 domains with Rd8_149 domains) and (13F replacement of Rd8_117 domains with Diverse_7 domains). A ribbon diagram of FnCas9 guide RNA shows circular rerouting of ωRNAs, which provides additional opportunities for engineering. (FIG. 8). IscB polypeptide Rd8_117 comprises a C-terminal helix, shown interacting with nucleotide strand in FIG. 14, another location for potential engineering of the IscB polypeptides.

Trans-RNA On-Switch and Off-Switch

[0945] The pseudoknot (PK) region of ωRNA can be engineered as a transRNA off-switch. (FIG. 15) As detailed in the application, it is possible to alter base identity of bases in the PK region to engineer a trans RNA off-switch. Similarly, the terminal hairpin region of ωRNA can be engineered as a transRNA on-switch wherein one could modify the structure of the ωRNA, with a portion of the ωRNA supplied in trans as the hairpin is needed for function of the IscB composition. (FIG. 16).

Activity of IscB with Cas9 REC domains

[0946] Applicants confirmed that chimeric IscBs with Cas9 REC domains are functional in vitro. (FIG. 17) Additionally, chimeric IscBs with Cas9 REC domains to extend enforced guide:target duplex are functional in vitro (FIG. 18). The chimeric IscBs with Cas9 REC domains mediate indel formation in human cells (Rd8_117) with FIG. 19 charting % indels for protein/REC insertion for guide 1 and guide 2. REC insertions on the IscB polypeptides destabilize TAM-distal guide mismatches with results of guides with varying mismatches shown in FIG. 20B. A further insertion utilizing 1261 REC in an IscB polypeptide accesses more target sites and improves indel activity in general at varying loci (FIG. 21).

IscB Polypeptide Engineering and TAM specificity

[0947] Applicants also investigated how TAM changed with insertion or exchange of various domains/subdomains. FIG. 22A shows weblogos indicating no changes in TAM of example Rd8_117 IscB polypeptide with insertion of full TI domain. The Rd8_117 IscB polypeptide with partial TI domain insertion showed no changes in TAM (FIG. 22B). Rd8_117 IscB polypeptide with Tudor domain insertions or Lance domain insertions shows no changes in TAM (FIG. 22C, 22D) However, only some Lance TIL domain insertions retain same TAM in example Rd8_117 IscB chimeric polypeptide (FIG. 22E); an insertion of Lance_TIL domain can lead to alterations in TAM of example Rd8_117 IscB polypeptide (FIG. 22F). A weblogo of altered TAMs in comparison to TAM from wild-type Rd_117 IscB polypeptide is shown in FIG. 23.

ωRNA engineering in IscB compositions

[0948] ωRNA engineering with Rd8_117 IscB polypeptide shows Rd8_117 can tolerate up to 21 nucleotides truncated from the 3' end of the ωRNA (FIG. 24A). However, insertions in the nexus pseudoknot of the ωRNA are minimally functional (FIG. 24B-24C). FIG. 24D shows that reduced activity from insertions in the nexus pseudoknot can be rescued by compensating insertions in nexus stem.

Example insertions and polypeptides are provided in Table 2, below.

Table 2.

SEQ ID NO:	Description
25194	Rd8_66_SpyCas9_Hyb_Stab_ins1 Hyb stabilizer domain from SpyCas9 comprising amino acids 309 to 327 was inserted starting at amino acid 308 of Rd8_117 polypeptide from Table 1
25195	Rd8_117_2089_REC_ins1 Rec domain from 2089 comprising amino acids 155 to 190, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 191 to 213 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1

25196	Rd8_117_Ace_REC_ins1 Rec domain 1 from AceCas9 comprising amino acids 155 to 304, Linker 1 comprising amino acids 305 to 307, Rec domain 2 from AceCas9 comprising amino acids 308 to 354, Linker 2 comprising 355 to 364, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising amino-acids 365 to 387 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25197	Rd8_117_Cas9_665_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 298, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 299 to 321 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25198	Rd8_117_Cas9_971_1_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 313, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 314 to 336 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25199	Rd8_117_Cas9_971_2_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 318, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 319 to 341 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25200	Rd8_117_Cas9_1079_1_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 259, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 260 to 282 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25201	Rd8_117_Cas9_1079_2_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 258, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 259 to 281 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25202	Rd8_117_Cas9_1079_3_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 265, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 266 to 288 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25203	Rd8_117_Cas9_1079_4_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 259, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 260 to 282 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25204	Rd8_117_Cas9_1261_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 301, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 302 to 324 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25205	Rd8_117_Cdi_REC_ins1 Rec domain 1 from CdiCas9 comprising amino acids 155 to 292, Linker 1 comprising amino acids 293 to 295, Rec domain 2 from CdiCas9 comprising amino acids 296 to 343, Linker 2 comprising 344 to 353, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising amino-acids 354 to 376 was inserted starting at amino acid 140of Rd8_117 polypeptide from Table 1
25206	Rd8_117_Cje_REC_ins1 Rec domain 1 from CjiCas9 comprising amino acids 155 to 276, Linker 1 comprising amino acids 277 to 279, Rec domain 2 from CjiCas9 comprising amino acids 280 to 323, Linker 2 comprising 324 to 333, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising amino-acids 334 to 356 was inserted starting at amino acid 140of Rd8_117 polypeptide from Table 1
25207	Rd8_117_Fno_REC_ins1 Rec domain from Cas9 comprising amino acids 155 to 301, Junction 1 comprising amino acids 141 to 154, and Junction 2 comprising 302 to 324 was inserted starting at amino acid 140 of Rd8_117 polypeptide from Table 1
25208	Rd8_117_Fno_REC_ins2
25209	Rd8_117_Fno_REC_ins3
25210	Rd8_117_Nme1_HNH_swap1
25211	Rd8_117_Nme1_REC_ins1
25212	Rd8_117_Nme1_RuvCIII_ins1
25213	Rd8_117_Rd8_66_REC_ins1
25214	Rd8_117_Sau_REC_ins1
25215	Rd8_117_Spy_REC_ins1
25216	Rd8_117_SpyCas9_Hyb_Stab_ins1

25217	Rd8_117_Sth_HNH_swap1
25218	Rd8_117_Sth_REC_ins1
25219	Rd8_117_REC_665_TI
25220	Rd8_117_REC_1079_1_TI
25221	Rd8_117_REC_1079_2_TI
25222	Rd8_117_REC_1079_3_TI
25223	Rd8_117_REC_2089_TI
25224	Rd8_117_REC_cA2_TI
25225	Rd8_117_REC_Diverse_7_TI
25226	Rd8_117_REC_Rd8_149_TI
25227	Rd8_117_REC_Cas9_665_tudor_lance
25228	Rd8_117_REC_Cas9_971_tudor
25229	Rd8_117_REC_Cas9_1079_1_tudor_lance
25230	Rd8_117_REC_Cas9_1079_1_tudor
25231	Rd8_117_REC_Cas9_1079_2_tudor_lance
25232	Rd8_117_REC_Cas9_1079_2_tudor
25233	Rd8_117_REC_Cas9_1079_3_tudor_lance
25234	Rd8_117_REC_Cas9_1079_3_tudor
25235	Rd8_117_REC_Cas9_1261_tudor_lance
25236	Rd8_117_REC_Cas9_1261_tudor
25237	Rd8_117_REC_ChlorIscB_tudor_lance
25238	Rd8_117_REC_ChlorIscB_tudor
25239	Rd8_117_REC_Diverse_7_NTS
25240	Rd8_117_REC_Diverse_7_REC
25241	Rd8_117_REC_Diverse_7_TI_L_tudor_lance
25242	Rd8_117_REC_Diverse_7_TI_L
25243	Rd8_117_REC_Diverse_7_TI_L1
25244	Rd8_117_REC_Diverse_7_TI_L2
25245	Rd8_117_REC_Diverse_7_TI_restore_WED
25246	Rd8_117_REC_Rd8_127_tudor_lance
25247	Rd8_117_REC_Rd8_149_NTS_1
25248	Rd8_117_REC_Rd8_149_NTS_2
25249	Rd8_117_REC_Rd8_149_TI_L
25250	Rd8_117_REC_Rd8_149_TI_restore_WED
25251	Rd8_117_REC_Rd8_149_tudor_nontudor
25252	Rd8_117_REC_Rd8_149_WED_ins
25253	Rd8_117_REC_Rd8_149_WED_swap
25254	Rd8_117_REC_644_2_tudor_ext
25255	Rd8_117_REC_644_2_tudor_lance
25256	Rd8_117_REC_644_2_tudor
25257	Rd8_117_REC_2089_tudor_lance
25258	Rd8_117_REC_2089_tudor
25259	Rd8_117_REC_cA2_tudor_lance
25260	Rd8_117_REC_cA2_tudor_swap
25261	Rd8_117_REC_Cas9_665_tudor
25262	Rd8_117_REC_SpCas9_RuvC_loop
25263	iscb_diverse_7-iscb_rd4_7-hrna-in-pab0001-bpii-backbone
25264	iscb_diverse_7-iscb_rd4_7-in-phs0728-pcdna3-1-cmv-sv40-nls-bamhi-nucleoplasmin-nls-3xha
25265	iscb_pathogens_14-quisiscb-hrna-pab0001-bpii-backbone
25266	iscb_pathogens_14-quisiscb-in-phs0728-pcdna3-1-cmv-sv40-nls-bamhi-nucleoplasmin-nls-3xha
25267	iscb-rd8_149-omegarna-in-pab0001
25268	phs0902-iscb_eukaryotic_1-ite-in-phs0728
25269	phs0928-eukaryotic-1-hrna-pab0001-b06-crrna-bpii-backbone
25270	phs1085-ii-d_1-sgrna-in-pab0001-bpii-backbone
25271	phs1087-ii-d_5-alt-sgrna-in-pab0001-bpii-backbone
25272	phs1089-ii-d_7-alt-sgrna-in-pab0001-bpii-backbone
25273	phs1107-ii-d_1-protein-in-phs0728-pcdna3-1-cmv-sv40-nls-bamhi-nucleoplasmin-nls-3xha
25274	phs1108-ii-d_5-protein-in-phs0728-pcdna3-1-cmv-sv40-nls-bamhi-nucleoplasmin-nls-3xha

25275	rd8_149-in-phs0728-pcdna3-1-cmv-sv40-nls-bamhi-nucleoplasmin-nls-3xha

Example 2. Characterization of Rd8_117 IscB

[0949] Applicants took biochemical measurements of Rd8_117, Rd8_66 and the chimeric Rd8_117 +Cas9 1079_1 REC insertion, with varying RNA:protein molar ratio from 4:1 to 0.25:1 (FIG. 96A). 1 uM protein was used in a buffer solution of 20 mM Tris ph 7.5, 5 mM MgCL2 and 50 mM NaCl with varying amounts of protein shows in vitro cleavage improves up to 4:1 molar ratio of RNA:protein. Varying temperature with measurement of activity of IscBs indicate Rd*_66 and Rd8_117 are meso-/thermophilic. RNA modifications also show modulation of activity at different temperatures using example chimeric IscB Rd8_117 with 1079_1 REC insertion (FIG. 96C). Cleavage kinetics were measured using 1 uM protein in the same buffer, with cleavage kinetics measured of the example chimeric IscB with REC insert. (FIG. 96D) Rd8_117 with REC inserts, Rd8_117 +1079_1 REC insert or Rd8_117 +1079_2 REC insert are functionally delivered by engineered Virus Like Particles (eVLPs) to HEK293 cells, with percent indels measured by 0. 0.005. 0.01. 0.05. 0.2. 1. 2 or 5 μL of eVLP delivered. (FIG. 97). Rd8_117+1079_1 REC is ~1/10 as efficient as plasmid transfection with highest dose.

Example 3. Chimeric IscB Base Editing

[0950] Applicants conducted an initial test of adenine ABE8 base editors with Rd8_117 IscBs, showing compatibility with ABE8 base editors. (FIG. 98A, 98B). Two guides were tested at *CTNNB1* and *CA2*, and one guide was tested at *DYNC1H1* and *HPRT1*, with percent editing measured at A positions noted on x-axis. (FIG. 98B). Chimeric IscB with 1079_1 REC insert shows improved adenine base editing at the same loci tested for the unmodified IscB ABE8 base editor. (FIG. 98C). The editing window for IscB adenine base editing is wide compared to Cas9 editing outcomes. Without being bound by theory it may be that the non-targeting sequence in the R loop is more exposed in the IscB protein than in Cas9.

Example 4. ωRNA modifications

[0951] Applicants undertook study of ωRNA modifications in example IscB polypeptides. Initial Rd8_66 and Rd8_117 demonstrate detectable binding of DNA targeting of IPTG1 via PAM -SCANR (FIG. 99A); with Rd8_66 indel formation appears affected by the inclusion of different structures of the ωRNA at VEGFA1 loci, including hairpin 3, hairpin 2+3, 3' truncation, hairpin 3 + 3' truncation and hairpin 2+3 + 3' truncation studied (FIG. 99B). Using Rd8_66, a 3' 20bp truncation was used to measure indel rates measurement across guides.

(FIG. 100). Figure 101 includes a schematic of engineered ωRNA that can be developed for trans on-switch and trans off-switch.

[0952] Rd8_117 cleavage activity measured with ωRNA 3' truncations indicated ωRNA tolerates up to 21 bp 3' truncations (FIG. 102A, 102B) but not 5' truncations (FIG. 102C, 102D).

[0953] Rd8_117 nexus loop reprogramming was explored (FIG. 103A), with 5'FAM/3'Cy5 labeled target, nexus loop modifications shown in FIG. 103B. Nexus pseudoknot insertion variants relative to WT ωRNA are indicated in FIG. 104A, with the cleavage activity of ωRNA variants shown in FIG. 104B. Further depictions of ωRNA folding and effects of 21 bp 3' truncation and relative to full ωRNA with Gibbs free energy measurements are indicated in FIG. 105.

[0954] The addition of the last 35 base pairs of Rd8_117 ωRNA can restore activity to the Rd8_117 35 bp 3' truncated ωRNA (FIG. 106A). FIG. 106B indicates 35 bp transRNA is active with multiple guides using example IscB Rd8_117 protein or Rd8_117 +1079 REC insertion chimeric IscB. Indications of -21 or -35 guide is a 3' truncation of 21 bp or 35 bp; similarly indications of -21 or -35 transRNA is a template consisting of last 3' 21 bp or 35 bp. Images from Cy5 channelRd8_117 ωRNA scaffold structure with truncations indicated. Trans-on RNA was shown to restore activity in mammalian cells, (FIG. 107A). Construct type measured for % indel indicated as follows: WT is the unmodified Rd8_117 IscB and 1079 indicates for results of Rd8_117 IscB with Cas 9 1079_1 REC insertion. A full transRNA= last 35 bp of Rd8_117 ωRNA scaffold; +G is addition of G after U6 and before the transRNA sequence; mini transRNA = 14 bp that span distance from -35 from 3' end to -21 from 3' end. Measurements made for DYNCH1 (107A) and HPRT1 (B).

Example 5. Chimeric IscB Mutations

[0955] Applicants tested mutations and measured activity improvement relative to Rd8_117 + Cas9_1079_1_REC insertion without mutations. The chimeric IscB polypeptide subjected to further mutations for gain of function evaluation has the following sequence:

Rd8_117_Cas9_1079_1 :

MRYVYVLDVDGKPLMPTCRFGKVRRMLKSGQAKAVDTLPFTIQLTYKPRTRILQPVTLGQDPGRTNIG
MAAVRFDGKELGRFHCITRNKEIPKLMADRMAARKASRRGERLARKRLARKLHTTAKHLNGRILPGC
SEPIAVKDIINTESRFNNRILTKCKVCGKNTPLRRNVRELLLENIVRFLPLESELKETLKRTILEGQQGNIN
KLFRKLKFNQKDWPGKNLTDIAKNKLPGRLPFCKEHFAENEKFTTIEKSTFRLTPTATQLLRTHINLFRK
LSGILPVTDVAVELNKFAFMQLDNPEMKKREIDFCHGPLCGTGGLEAAVKEQQDGKCLLCGKESIGHY

HHIVPRSRRGSNTIANIAGLCPKCHELVHKDADTAESLTEMKTGLMKKYGGTSVLNQIIPKLVETLADL FPGHFHVTNGWNTKEFREKHHLEKDHDVDAYCIACSHLKPEETLVETEPFEILQFRKHNRAIIHHQTER TYKLDGVTVAKNRKKRMEQKTDSLEDWYVDMAKEHGKTQADAMRSRLTVIKSTRYYNTPGRMMPG TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKVRILTKNTGLVFVA (SEQ ID NO: 25278).

[0956] Mutations can be selected by investigating the natural distribution of amino acids at each position and sampling the distributions for desired changes. For example, if additional binding to DNA is desired in different regions, K, R, or H substitutions were identified by identifying positions that naturally tolerate K, R, H based on a multiple sequence alignment. Figures 108-115 show the different mutations in affecting various allosteric mechanisms involved with DNA binding and protein packing as well as indicating fold improvement in IscB activity with particular mutations.

Table 3. Mutations in Rd8_117 Cas9 1079_1 REC insertion and Activity Fold Improvement Averaged over two guides

MUTATION	FOLD INCREASE ACTIVITY	Notes
Q60I	1.7435468168893,	hydrophobic repacking of RuvC
A288V	2.07450864829255	hydrophobic repacking of RuvC
G395R	1.1227821869606	
W425R	1.64932718864797	hydrophobic repacking of RuvC
T358R	1.16183142694273	enhancing existing DNA/RNA channel
S325R	2.07347940926662	
M500K	1.27873398239969	
E409R	2.17304365761619	extending the DNA/RNA duplex channel
D413K	2.22701849334025	extending the DNA/RNA duplex channel
P405R	1.84352340394802	extending the DNA/RNA duplex channel
E137K	1.69421379430915	enhancing existing DNA/RNA channel
T147R	1.03058538524008	
L260K	1.2254347524772	
E308R	1.47748789689566	extending the DNA/RNA duplex channel
L467R	1.92783745258766	
N66H	1.6802402145208	new non-targeting strand channel
T65K	1.35060670006904	
I85E	1.81117812402217	
I85R	1.85091381825009	new non-targeting strand channel
E90R	1.20650321021699	

E501K	0.084486229388418	enhancing existing DNA/RNA channel
M500K	1.09694754180498	enhancing existing DNA/RNA channel
S352H	1.74258630946715	enhancing existing DNA/RNA channel
K93D	1.35540572541723	
T561R	1.85242173829211	altering unwinding activity
E576R	2.04892696341647	altering unwinding activity
G417K	1.9385824020467	extending the DNA/RNA duplex channel
L291R	2.12332880739941	extending the DNA/RNA duplex channel
V490K	1.96187720091516	non-specific DNA binding mutations
V490R	2.04413299842519	non-specific DNA binding mutations
G489K	1.8661312489374	non-specific DNA binding mutations
T481R	1.76169099892367	non-specific DNA binding mutations
T536R	2.0215905621095	non-specific DNA binding mutations
I533K	2.22498173347667	non-specific DNA binding mutations
T484K	2.04784709609413	non-specific DNA binding mutations
N566R	1.83160206355917	non-specific DNA binding mutations
Y538R	1.42813000040608	non-specific DNA binding mutations
H271I	1.00520784107492	hydrophobic repacking of RuvC

[0957] Applicants are further exploring combination mutations that consider limited improving mutations from each of the major allosteric mechanisms involved in the cleavage process. Applicants plan on improving activity in multiple distinct networks, aiming to achieve a better level of independence between the mutations, thereby increasing their combinability. Table 4 below provides examples of mutations that will be tested, relative to Rd8_117_Cas9_1079_1.

Table 4. Combinations of Mutations

Description	**Mutations**					
Pack	Q60I	W425R	A288V			
Pack_del1		W425R	A288V			
Pack_del2	Q60I		A288V			
Pack_del3	Q60I	W425R				
Distal	P405R	E409R	E308R	L291R		

Distal_del1		E409R	E308R	L291R		
Distal_del2	P405R		E308R	L291R		
Distal_del3	P405R	E409R		L291R		
Distal_del4	P405R	E409R	E308R			
Blend	I85R	E576R	E137K	I533K	E409R	A88V
Blend_del1		E576R	E137K	I533K	E409R	A88V
Blend_del2	I85R		E137K	I533K	E409R	A88V
Blend_del3	I85R	E576R		I533K	E409R	A88V
Blend_del4	I85R	E576R	E137K		E409R	A88V
Blend_del5	I85R	E576R	E137K	I533K		A88V
Blend_del6	I85R	E576R	E137K	I533K	E409R	
P1	E409R	E576R				
P2	E409R	A288V				
P3	E409R	I533K				
P4	E409R	I85R				
P5	E409R	E137K				
P6	P405R	E576R				
Unwind	T561R	E576R				
Stabilize	E137K	S352H				

Example 6. Natural Diversity-Inspired Engineering of OrufIscB for High Efficiency Genome Editing

[0958] Applicants sought to increase the effective guide length by introducing REC domains from Cas9s into OrufIscB. SpCas9 has been shown to effectively use the first 20 bp of the guide sequence[266], which is also protected by the protein when bound to target DNA[261,262]. In contrast, superimposing the AlphaFold2 structure of OrufIscB onto cryo-EM structures of the OgeuIscB ternary complex[261,262] suggests that only about 16 bp of the guide:target heteroduplex are protected by IscB. Applicants therefore reasoned that increasing the surface area of protein that could interact with the guide:target heteroduplex could increase the effective guide length. Additionally, the Cas9 REC domain, in particular the REC2 and REC3 subdomains, has been shown to be involved in PAM-distal guide:target heteroduplex mismatch detection in SpCas9[264,265] in addition to target unwinding and interaction. Applicants hypothesized that an inserted REC domain may confer mismatch detection to an extended guide as well.

[0959] To determine where to insert REC domains in the OrufIscB scaffold, Applicants performed a multiple sequence alignment of several IscB and Cas9 proteins, and observed

homology at sites flanking the REC domain of Cas9s and REC-like insert of IscBs, in particular between REC-containing IscBs and the recently identified Type II-D Cas9s[215], which are the closest relatives of IscB (FIG. 116). Notably, despite previous observations that the REC domain of Cas9 is generally not conserved[268–274], Applicants found that the small REC domains in Type II-D Cas9 REC domains share several particularly conserved residues (FIG. 116), which are also shared by TbaIscB, the founding member of the Cas9 family[215], and CzcbIscB, suggesting a common origin of REC domains despite extant diversity. Given the homology of the sites surrounding REC domains, Applicants reasoned that Cas9 REC domains could be inserted into the homologous location in OrufIscB. Applicants constructed OrufIscB-REC domain chimeras with 183 REC domains derived largely from the Type II-D Cas9s, as well as several from Cas9s commonly used for genome editing such as SpCas9 and SaCas9, and some IscBs with divergent REC-like insertions.

[0960] To assess the ability of these REC domain insertions to increase guide length and retain high dsDNA cleavage efficiency, Applicants tested each variant using an IVTT target/guide screen, in which Applicants included a library of 100 ωRNAs targeting human genomic sequences and a corresponding library of targets including every possible mismatch and increasing numbers of TAM-distal mismatches to assess the effective guide length (FIG. 117A). Applicants measured the normalized read count of cleaved products derived from each variant in the target library and compared to the wild-type OrufIscB. From this screen, Applicants identified 54 REC insertions that exhibited decreased cleavage with 5, 6 and 7 TAM-distal mismatches relative to wild-type OrufIscB, while still maintaining appreciable on-target efficiency (see Methods) (FIG. 117B). Applicants selected these candidates for further validation in human cells and found that several improved the on-target activity of OrufIscB (FIG. 118). In particular, Applicants identified the REC domain from NbaCas9-1, which displayed the best improvement on average across 6 tested guides (FIG. 118). Applicants further compared the activity of wild-type OrufIscB and OrufIscB with the NbaCas9-1 REC domain and found that the chimeric protein improved activity up to 20-fold for some guides, as well as displayed activity at several sites that could not be targeted by the wild-type protein (FIG. 117C). Applicants assessed the optimal guide length of the chimeric OrufIscB with the NbaCas9-1 REC domain insertion at 4 target sites in the human genome and compared the resulting distribution with wild-type OrufIscB. Applicants found that the chimeric version exhibited optimal activity with ~19-20 bp guides (FIG. 119), suggesting that the insertion of

the NbaCas9-1 REC domain skewed the optimal guide length to be longer. Applicants therefore continued with the OrufIscB including the NbaCas9-1 REC insertion, which Applicants termed enOrufIscB-v1, for further optimization.

[0961] Based on natural variation at homologous sites across the diversity of IscBs at positions likely contacting target DNA or ωRNA, Applicants selected 38 amino acid changes and tested them on top of enOrufIscB-v1. Applicants found that most variants improved activity relative to enOrufIscB-v1 across 2 tested guides (FIG. 117D), illustrating the utility of natural variant-guided rational engineering. From these, Applicants selected 3 mutants at different locations in the predicted structure of OrufIscB, as Applicants hypothesized that improvements to DNA binding kinetics by different domains may act synergistically to facilitate enhanced DNA interaction (FIG. 117E). Applicants tested all possible combinations of these 3 mutants, including single, double and triple mutants, at a panel of 12 sites and found that the triple mutant exhibited the best improvement in activity relative to enOrufIscB-v1 and exhibited indel rates up to 45% for some sites (FIG. 117F). Applicants therefore termed this enOrufIscB-v2.

Application of enOrufIscB-v2 as a versatile genome interrogation tool

[0962] Although genome editing has traditionally relied upon induction of double-stranded breaks (DSBs) and repair of break sites to generate genomic edits, DSBs can cause adverse effects such as chromosomal translocations[266] and growth arrest due to activation of p53-mediated DNA damage response[258]. Therefore, genome modification platforms such as base editing245, prime editing[246], and epigenome editing[76,78,217,275] that do not require DSBs, are promising therapeutic modalities for avoiding such undesired outcomes. As both base editing and transcriptional repression systems have already demonstrated positive preclinical results[76,78,217,276] and are desirable for conducting in vivo screens, Applicants sought to further develop these platforms for in vivo application using enOrufIscB-v2.

[0963] Given its compact size, enOrufIscB-v2 fused to the required effector domains for base editing and transcriptional repression along with the corresponding ωRNA can be packaged in a single AAV genome for therapeutic delivery (FIG. 120A). Applicants set out to develop this platform by first assessing the compatibility of enOrufIscB-v2 with these genome editing modalities. Applicants began by fusing enOrufIscB-v2 with base editing enzymes using off-the-shelf construct designs optimized for SpCas9. For adenine base editing, Applicants used ABE8e[276], which was optimized using phage-assisted continuous evolution. Applicants

transfected this construct in HEK293FT cells along with 10 guides containing adenines at a variety of positions, to simultaneously assess the editing window for each base editor. Applicants found that enOrufIscB-v2-ABE was capable of efficient base editing at all sites tested with varying efficiencies roughly corresponding to the efficiency of indel generation at the same sites by cleavage-competent enOrufIscB-v2 (FIG. 120B). The editing window was approximately at positions 3-17, counted from the 5′ end of the guide sequence.

[0964] Applicants then developed a transcriptional repression system using enOrufIscB-v2. Applicants fused enOrufIscB-v2 with the DNMT3A and DNMT3L DNA methylation writers and a KRAB transcriptional repression domain, again using an off-the-shelf design originally optimized for SpCas9 (CRISPRoff-v2.1)[277], and termed this system OMEGAoff. Applicants transfected OMEGAoff in HEK293FT cells with guides targeting the promoter regions of several clinically relevant genes, such as FABP4, using previously determined rules for relative positioning to the transcription start site[76,88,89,258]. Applicants found that OMEGAoff was capable of repressing transcription of all targeted genes to similar levels as CRISPRoff-v2.1 as measured by qPCR (FIG. 120C). Together, these results demonstrate the flexibility of enOrufIscB-v2 to mediate diverse genome interrogation modalities. Further, the mechanistic similarities between IscB and Cas9 enable straightforward construct design using systems previously optimized for Cas9.

Materials and methods associated with Sections 3.8-3.11

Cell-free transcription/translation TAM interference assay

[0965] IscB protein sequences were human codon optimized using the GenScript codon optimization tool. IscB genes and ωRNA scaffolds were custom synthesized by Twist Biosciences, and transcription/translation templates were generated by PCR from custom synthesis products. Cell- free transcription/translation reactions were carried out using a PURExpress In Vitro Protein Synthesis Kit (NEB) as per the manufacturer's protocol with half-volume reactions, using 75 ng of template for the protein of interest, 125 ng of template for the corresponding ωRNA with a guide targeting the TAM library and 25 ng of TAM library plasmid. Reactions were incubated at 37°C for 4 hours, then quenched by placing at 4°C or on ice and adding 10 μg RNase A (Qiagen) and 8 units Proteinase K (NEB) each followed by a 5 min incubation at 37°C. DNA was extracted by PCR purification and adaptors were ligated using an NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) using the NEBNext Adaptor for Illumina (NEB) as per the manufacturer's protocol. Following adaptor ligation,

cleaved products were amplified specifically using one primer specific to the TAM library backbone and one primer specific to the NEBNext adaptor with a 12-cycle PCR using NEBNext High Fidelity 2X PCR Master Mix (NEB) with an annealing temperature of 63°C, followed by a second 18-cycle round of PCR to further add the Illumina i5 adaptor. Amplified libraries were gel extracted, quantified by qPCR using a KAPA Library Quantification Kit for Illumina (Roche) on a StepOne Plus machine (Applied Biosystems/Thermo Fisher Scientific) and subject to single-end sequencing on an Illumina MiSeq with Read 1 80 cycles, Index 1 8 cycles and Index 2 8 cycles. TAMs were extracted and enrichment score for each PAM was calculated by filtering for all TAMs present more than once and normalizing to the TAM frequency in the input library subject to the same in vitro transcription/translation and quenching reactions. A position weight matrix based on the enrichment score was generated and weblogos were visualized based on this position weight matrix using a custom Python script.

Mammalian cell culture and transfection

[0966] Mammalian cell culture experiments were performed in the HEK293FT line (Thermo Fisher) grown in Dulbecco's Modified Eagle Medium with high glucose, sodium pyruvate, and GlutaMAX (Thermo Fisher), additionally supplemented with 1× penicillin–streptomycin (Thermo Fisher), 10 mM HEPES (Thermo Fisher), and 10% fetal bovine serum (VWR Seradigm). All cells were maintained at confluency below 80%.

[0967] All transfections were performed with Lipofectamine 3000 (Thermo Fisher). Cells were plated 16-20 hours prior to transfection to ensure 90% confluency at the time of transfection. For 96-well plates, cells were plated at 20,000 cells/well, and for 24-well plates, cells were plated at 100,000 cells/well. For each well on the plate, transfection plasmids were combined with 2 μL of P3000 solution per every 1 μg DNA and OptiMEM I Reduced Serum Medium (Thermo Fisher) to a total of 25 μL. Separately, 23 μL of OptiMEM was combined with 2 μL of Lipofectamine 2000. Plasmid and Lipofectamine solutions were then combined and pipetted onto cells.

Mammalian genome editing

[0968] ωRNA scaffold backbones were cloned into a pUC19-based human U6 expression backbone by Gibson Assembly. Human codon-optimized IscB genes were cloned into a CMV expression backbone by Gibson assembly using 2X Gibson Assembly Master Mix (NEB) to generate pCMV-SV40 NLS-IscB protein-nucleoplasmin NLS-3xHA constructs. For initial

testing, 12-guide libraries were cloned in a pool mixing primers to add each of the 12 guides in in a given pool at equimolar ratios, and ωRNA scaffold backbones were subject to whole plasmid amplification with guide primers annealing to the U6 promoter and a second primer annealing to the start of the ωRNA scaffold using Phusion Flash High-Fidelity 2X Master Mix (Thermo Fisher Scientific). PCR products were gel extracted and eluted in 30 uL, then blunt-end ligated to circularize by addition of 5 units T4 PNK (NEB), 200 units T4 DNA Ligase (NEB) and final 1X T4 DNA Ligase Buffer (NEB) and incubation for 1.5 h at room temperature prior to transformation in Stbl3 chemically competent E. coli (NEB). For individual guide constructs, oligos with appropriate overhangs were synthesized by Genewiz, annealed and phosphorylated using T4 PNK (NEB) and cloned into ωRNA backbones by restriction-ligation cloning. Human codon-optimized IscB genes were cloned into a CMV expression backbone by Gibson assembly using 2X Gibson Assembly Master Mix (NEB) to generate pCMV-SV40 NLS-IscB protein- nucleoplasmin NLS-3xHA constructs.

[0969] Prior to testing individual guides, each tested IscB protein was screened for activity in HEK293FT using a pool of 12 guides cloned as described. For this 12-guide pooled initial screening of IscB proteins, 800 ng protein expression construct and 800-1200 ng of the corresponding guide pool with corresponding ωRNA scaffold were transfected in one well of a 24-well plate as described. After 60-72 hours, genomic DNA was harvested by washing the cells once in 1xDPBS (Sigma Aldrich) and dry trypsinizing cells using TrypLE (Thermo Fisher Scientific). Trypsinized cells were collected in 1 mL 1xDPBS and pelleted by centrifugation at 300xg at 4°C for 5 min. The supernatant was removed and cells were resuspended in 50 uL QuickExtract DNA Extraction Solution (Lucigen) and cycled at 65°C for 15 min, 68°C for 15 min then 95°C for 10 min to lyse cells. 2.5 μL of lysed cells were used as input into each PCR reaction. Amplification of each region targeted by a guide in a given guide pool was performed individually.

[0970] For individual guide sequences, 100 ng guide/ωRNA expression plasmid and 100 ng protein expression plasmid were transfected in each of 4 wells as biological replicates in a 96- well plate for each guide condition as described. After 60-72 hours, genomic DNA was harvested by washing the cells once in 1xDPBS (Sigma Aldrich) and adding 50 uL QuickExtract DNA Extraction Solution (Lucigen). Cells were scraped from the plates to suspend in QuickExtract and cycled at 65°C for 15 min, 68°C for 15 min then 95°C for 10 min to lyse cells. 2.5 μL of lysed cells were used as input into each PCR reaction.

[0971] For library amplification, target genomic regions were amplified and with a 12-cycle PCR using NEBNext High Fidelity 2X PCR Master Mix (NEB) with an annealing temperature of 63°C for 15 s, followed by a second 18-cycle round of PCR to add Illumina adapters and barcodes. The libraries were gel extracted and subject to single-end sequencing on an Illumina MiSeq with Read 1 300 cycles, Index 1 8 cycles, Index 2 8 cycles. Insertion/deletion (indel) frequency was analyzed using CRISPResso2[212], with a quantification window center of -9 and a window size of 6 based on a previous analysis of IscB cleavage patterns[215]. To eliminate noise from PCR and sequencing error only indels with at least 2 reads or more than 1 base inserted or deleted were counted towards reported indel frequencies. For 12-guide pooled screens, read alignments were further inspected manually for presence of "true" indels to select candidates for validation. For individual guide/ωRNA experiments, to assess statistical significance, 2-tailed T- tests were performed using non-targeting guide/ωRNA conditions as a negative control.

Cell free transcription-translation mismatch tolerance assay

[0972] The target library was designed by selecting 100 random sites from coding sequences in the human genome (hg38 assembly) adjacent to ATAAA TAMs. Target sites were selected to avoid homopolymeric sequences of 4 or more Ts or Gs to avoid guides with potential termination signals for later use with PolIII promoters and those with potential to form G quadruplexes. Targets were also selected to have an edit distance of at least 3 away from all other targets in the TAM-proximal 7 bp, and at least edit distance of 10 in total. All possible single mismatches for each target were then generated, and progressive mismatch targets were also generated by selecting 8 sequences each with TAM-distal mismatches ranging from 2 to 7 bp at the 5′ end of the target, with at least an edit distance of 8 from any of the non-template original targets in the library. 316 random sequences with an edit distance of at least 10 from all other library members were added as negative controls. 10N barcodes were added to each individual target for distinguishing target identity after cleavage. The library was synthesized by GenScript and cloned into a pUC19 vector by Gibson Assembly.

[0973] An associated guide library was synthesized by IDT and cloned into a backbone containing the OrufIscB ωRNA scaffold by Gibson Assembly. In vitro transcription templates for the pooled guide library were then generated by PCR and were transcribed using a HiScribe T7 Quick High Yield RNA synthesis kit (NEB). In vitro transcription/translation templates for IscB proteins with REC domain insertions were prepared as described in "Cell-free

transcription/translation TAM interference assay" above. Cell-free transcription/translation reactions were carried out using a PURExpress In Vitro Protein Synthesis Kit (NEB) as per the manufacturer's protocol with half-volume reactions, using 75 ng of template for the protein of interest, 1.5 uM final concentration of the in vitro-transcribed ωRNA guide library and 25 ng of target library plasmid. Cas9 and a guide targeting a non-library control were also included as an internal activity control for downstream library preparation and sequencing. Reactions were carried out and libraries were prepared as in "Cell-free transcription/translation TAM interference assay" above. Barcodes were extracted and the read count of all mismatch targets was normalized to the read count of the associated perfectly matched target to generate relative scores for cleavage of "off-target" substrates. Medians for various categories were then counted and analyzed using a custom Python script.

Mammalian base editing assays

[0974] Constructs expressing base editors and associated guideRNAs were transfected, gDNA was harvested and libraries were prepared as described for individual protein/guide combinations in "Mammalian cell culture and transfection" and "Mammalian genome editing" above. Editing was quantified by counting the number of reads at which the expected edited position in the amplicon was called as a G (on top strand) or C (on bottom strand) and dividing by the total number of reads in the sample using a previously described custom Python script[278]. Unless otherwise noted, all reported data is the average of 4 biological replicates.

Mammalian transcriptional repression assays

[0975] Constructs expressing base editors and associated guideRNAs were transfected as described for individual protein/guide combinations in "Mammalian cell culture and transfection" and "Mammalian genome editing" above. RNA was extracted using an RNeasy 96 Plus kit (Qiagen) and reverse transcribed using a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher) using random hexamer primers as previously described[278]. RNA expression was measured by qPCR using commercially available TaqMan probes (Thermo Fisher) on a LightCycler 480 II (Roche) with GAPDH as an endogenous internal control in 5 uL multiplexed reactions. Each of the 4 biological replicate is the average of 4 technical qPCR replicates, and relative expression was calculated using the ddCt method279 with a negative control condition consisting of the corresponding OMEGAoff or CRISPRoffv2.1 expression plasmid co-transfected with an AAVS targeting guide. Statistical significance was assessed using a two-tailed t-test.

Bibliography for Example 6

76. Yan, W. X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science* **363**, 88–91 (2019).

78. Shmakov, S. *et al.* Discovery and functional characterization of diverse Class 2 CRISPR- Cas systems. *Mol. Cell* **60**, 385–397 (2015).

88. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).

89. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. & Koonin, E. V. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5307–E5316 (2018).

212. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

215. Altae-Tran, H. *et al.* The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).

217. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).

246. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101–1108 (2008).

258. Kannan, S. *et al.* Compact RNA editors with small Cas13 proteins. *Nat. Biotechnol.* **40**, 194–197 (2022).

261. Kato, K. *et al.* Structure of the IscB-ωRNA ribonucleoprotein complex, the likely ancestor of CRISPR-Cas9. *Nat. Commun.* **13**, 6719 (2022).

262. Schuler, G., Hu, C. & Ke, A. Structural basis for RNA-guided DNA cleavage by IscB- ωRNA and mechanistic comparison with Cas9. *Science* **376**, 1476–1481 (2022).

264. Pacesa, M. *et al.* Structural basis for Cas9 off-target activity. *Cell* **185**, 4067-4081.e21 (2022).

265. Pacesa, M. *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature* **609**, 191–196 (2022).

266. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).

268. Musunuru, K. *et al.* In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).

269. Rothgangl, T. *et al.* In vivo adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. *Nat. Biotechnol.* **39**, 949–957 (2021).

270. Moreno, A. M. *et al.* Long-lasting analgesia via targeted in situ repression of NaV1.7 in mice. *Sci. Transl. Med.* **13**, eaay9056 (2021).

271. Choi, E. H. *et al.* In vivo base editing rescues cone photoreceptors in a mouse model of early-onset inherited retinal degeneration. *Nat. Commun.* **13**, 1830 (2022).

272. Su, J. *et al.* In vivo base editing rescues photoreceptors in a mouse model of retinitis pigmentosa. *Mol. Ther. Nucleic Acids* **31**, 596–609 (2023).

273. Backstrom, J. R., Sheng, J., Wang, M. C., Bernardo-Colón, A. & Rex, T. S. Optimization of S. aureus dCas9 and CRISPRi elements for a single adeno-associated virus that targets an endogenous gene. *Mol. Ther. Methods Clin. Dev.* **19**, 139–148 (2020).

274. Newby, G. A. & Liu, D. R. In vivo somatic cell base editing and prime editing. *Mol. Ther.* **29**, 3107–3124 (2021).

275. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).

276. Shmakov, S. A. *et al.* Systematic prediction of functionally linked genes in bacterial and archaeal genomes. *Nature Protocols* vol. 14 3013–3031 Preprint at https://doi.org/10.1038/s41596-019-0211-1 (2019).

277. Makarova, K. S. *et al.* Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).

278. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

***

[0976] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is:

1. An engineered chimeric IscB composition comprising:

a) an IscB polypeptide comprising one or more insertions of a heterologous polypeptide, and optionally one or more modified amino acids, that increases specificity or activity of the chimeric IscB polypeptide relative to wildtype; and

b) an ωRNA molecule comprising a scaffold and a reprogrammable spacer sequence, the ωRNA molecule capable of forming a complex with the IscB polypeptide and directing sequence-specific binding of the IscB polypeptide to a target polynucleotide.

2. The IscB composition of claim 1, wherein the insertion is a Rec domain, or functional fragment thereof.

3. The IscB composition of claim 2, wherein the Rec domain is from a Type II Cas polypeptide.

4. The IscB composition of claim 3, wherein the Rec domain is a Rec domain from a Type II-D Cas polypeptide.

5. The IscB composition of claim 3, wherein the Rec domain is a Rec domain from a Cas9.

6. The engineered IscB of claim 5, wherein the Cas9 is derived from *Francisella novicida* (FnoCas9), *Neisseria meningitidis* (NmeCas9), *Staphylococcus aureus* (SaCas9), *Streptococcus pyogenes* (SpCas9), *Streptococcus thermophilus* (StCas9), *Acidothermus cellulolyticus* (AceCas9), *Campylobacter jejuni* (CjeCas9) or a combination thereof.

7. The IscB composition of anyone of claims 2 to 6, wherein the Rec domain is inserted between amino acids 153-160 of RD8_117 IscB polypeptide of Table 1, or an analogous position of another IscB polypeptide.

8. The IscB composition of anyone of claims 2 to 7, wherein the ωRNA comprises a deletion that reduces steric interference with the inserted Rec domain.

9. The IscB composition of claim 8, wherein the deletion is in the PK-loop of the ωRNA.

10. The IscB composition of claim 9 wherein the deletion comprises 1 to 25, more preferably 1 to 21 nucleotides of Rd8_117 reference ωRNA or Table 1 or an analogous position in another ωRNA.

11. The IscB composition of claim 1, wherein the insertion is a protein capable of binding to RNA, DNA, or both.

12. The IscB composition of claim 11, wherein the insertion is a nuclease, or a functional fragment thereof.

13. The IscB composition of claim 12, wherein the insertion is an endonuclease, exonuclease, or functional fragment thereof.

14. The IscB composition of claim 13, wherein the endonuclease is a Ribonuclease (RNase), deoxyribonuclease (DNase), or fragment thereof.

15. The IscB composition of claim 11, wherein the insertion is hybrid binding domain (HBD).

16. The engineered IscB of claim 11, wherein the insertion is a RuvC domain or portion thereof.

17. The engineered IscB of claim 16, wherein the RuvC domain comprises a Cas9 RuvC domain/region, subdomain/subregion, or portion thereof.

18. The IscB composition of claim 1, wherein the insertion is a TAM interacting (TI) or PAM interacting (PI) domain, or functional fragment thereof.

19. The IscB composition of claim 18, wherein the domain is an NGG PI or TI domain, or functional fragment thereof.

20. The IscB composition of any of the preceding claims, wherein TAM determining region comprises one or more amino acid substitutions.

21. The IscB composition of claim 20, wherein the insertion is in the WED/adaptor stabilizer region, the Tudor domain, Tudor Lance domain, or a combination thereof.

22. The IscB composition of claim 20 or 21, wherein the insertion preserves RNA interaction with the TAM determining region.

23. The IscB composition of anyone of claims 18 to 22, wherein the TI domain, PI domain, or functional fragment thereof is inserted between amino acids 365-499 of Rd8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide.

24. The IscB composition of claim 18 to claim 23, wherein the TI domain, PI domain, or functional fragment thereof is from a Type II Cas polypeptide.

25. The IscB composition of claim 18 to claim 23, wherein the TI domain or functional fragment thereof is from an IscB polypeptide.

26. The IscB composition of claim 25, wherein the insertion replaces amino acids 365-499, 369-499, 462-486 (Lance), 450-486 (Tudor), or 376-383, 436-448, and 462-486 (Lance_TIL) of the TI domain of Rd8_117 polypeptide of Table 1, or an analogous position of another IscB polypeptide.

27. The IscB composition of claim 24, wherein the TAM of the wild-type IscB polypeptide is retained or wherein the TAM is modified.

28. The IscB composition of claim 23, wherein the insertion comprises one or more amino acid positions from 380-735 from SEQ ID NO: 2365 one or more amino acid positions 556-609 from cA2 ProCas9-2, one or more amino acid positions 356-420 from IscB_Rd8_149, one or more amino acid positions 386-464 from IscB_Rd4_7, one or more amino acid positions 856-924 from Cas9_971, one or more amino acid positions 569-751 from Cas9_1079_3, one

or more amino acid positions from ChlorIscB, one or more amino acid positions 488-512 from SEQ ID NO: 2367, one or more amino acid positions 739-765 from SEQ ID NO: 2365, one or more amino acid positions 407-431 from IscB_Rd8_127, one or more amino acid positions 404-430 from CRISPR IscB 00644, one or more amino acid positions 376-482 from IscB_large_28, one or more amino acid positions 356-488 from IscB_Rd8_149, one or more amino acid positions 376_482 from IscB_Rd8_75, one or more amino acid positions 374-495 from IscB_Rd8_151, one or more amino acid positions 376-477 from IscB_Rd8_23, one or more amino acid positions 376-477 from IscB_Rd8_24, one or more amino acid positions 376-486 from IscB_Rd8_118, one or more amino acid positions 374-495 from IscB_Rd8_151, and/or one or more amino acid positions 376-486 amino acids from IscB_Rd8_118.

29. The engineered IscB of any of the preceding claims, wherein one or more nucleotides in a pseudoknot nexus of the ωRNA is modified and/or inserted.

30. The engineered IscB of claim 29, wherein one or more nucleotides in a nexus stem of the ωRNA that base pair to the one or more nucleotides in the pseudoknot nexus is modified and/or inserted.

31. The engineered IscB of claim 30, wherein a base pair comprising a nucleotide in the pseudoknot nexus and a nucleotide in the nexus stem is substituted with a complementary base pair.

32. The engineered IscB of any of the preceding claims, wherein one or more nucleotides in a pseudoknot region of the ωRNA are modified and/or inserted and wherein the nexus stem retains base pairing.

33. The engineered IscB of any of the preceding claims, wherein one or more nucleotides in a pseudoknot region of the ωRNA are modified and/or inserted and wherein the pseudoknot retains its structure relative to a wild-type IscB.

34. The engineered IscB of claim 33, wherein the pseudoknot region of the ωRNA retains its structure by no less than 50%, no less than 55%, no less than 60%, no less than 65%, no less than 70%, no less than 75%, no less than 80%, no less than 85%, no less than 90%, no less than 95% relative to a wild-type IscB.

35. The engineered IscB of any of the preceding claims, wherein the pseudoknot comprises a peptide nucleic acid (PNA).

36. The engineered IscB of any of the preceding claims, wherein the 3'-end of the ωRNA is truncated.

37. The engineered IscB of claim 36, wherein the 3'-end of the ωRNA is truncated by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides.

38. The engineered IscB of claim any of the preceding claims, wherein a RNA supplied in trans is bound to a 3'-end of the ωRNA.

39. The engineered IscB of any of the preceding claims, wherein one or more amino acids in contact with a nucleotide are substituted or removed.

40. The engineered IscB of claim 39, wherein one or more amino acids in contact with the ωRNA are substituted or removed.

41. The engineered IscB of any of the preceding claims, wherein one or more amino acids on the surface of the engineered IscB are substituted or removed.

42. The engineered IscB of claim 41, wherein the one or more amino acids on the surface of the engineered IscB are substituted to have a different charge.

43. The engineered IscB of claim 41, wherein the one or more amino acids on the surface of the engineered IscB are removed or substituted to increase or decrease the overall charge of the protein surface.

44. The engineered IscB of any of the preceding claims, wherein the surface charge of the engineered IscB is more neutral relative to the wildtype.

45. The engineered IscB composition of any of the previous claims, further comprising a nucleotide deaminase.

46. The engineered IscB of claim 11, wherein the insertion is a nucleotide deaminase

47. The engineered IscB of claim 45 or 46, wherein the nucleotide deaminase is an adenosine deaminase or cytidine deaminase.

48. The engineered IscB of claim 45, wherein the nucleotide deaminase is inserted in the middle of the IscB or fused at the N terminus or C-terminus of the IscB polypeptide.

49. The engineered IscB of claim 45 to 48, wherein the cytosine deaminase is apolipoprotein B mRNA-editing enzyme, catalytic polypeptide (APOBEC), an activation-induced deaminase (AID), a cytidine deaminase 1 (CDA1), or cytosine deaminase acting on RNA (CDAR).

50. The engineered IscB of claim 47, wherein the adenosine deaminase is ADAR or TadA.

51. The engineered IscB of claim 50, wherein the adenosine deaminase is an ADAR and the reprogrammable spacer sequence of the ωRNA molecule comprises one or more mismatches to the target polynucleotide.

52. The IscB composition of any of the preceding claims, wherein the C-terminus is engineered for base-editing.

53. The engineered IscB of claim 52, wherein the insertion is a transposase.

54. The engineered IscB of claim 53, wherein the transposase is a TnpA.

55. The engineered IscB of any of the preceding claims, further comprising a functional domain.

56. The engineered IscB of claim 55, wherein the functional domain comprises a base editing system or fragment thereof.

57. The engineered IscB of claim 55, wherein the functional domain comprises a prime editing system or fragment thereof.

58. The engineered IscB of claim 55, wherein the functional domain comprises an epigenetic editing system or fragment thereof.

59. The engineered IscB of any one of claims 55-58, wherein the functional domain is inserted at a junction.

60. The engineered IscB of any one of claims 1 to 59, further comprising one or more mutations of Table 3 or Table 4.

61. A method of modifying a target nucleotide, optionally within a cell, comprising contacting the target nucleotide or the cell with the composition of any one of claims 1 to 60.

62. A vector system comprising one or more vectors encoding the IscB polypeptide and the ωRNA molecule of claim 1.

63. One or more polynucleotides encoding one or more components of the composition of any of claims 1 to 60.

64. One or more vectors encoding the one or more polynucleotides of claim 63.

65. An engineered cell comprising the composition of any of claims 1 to 60.

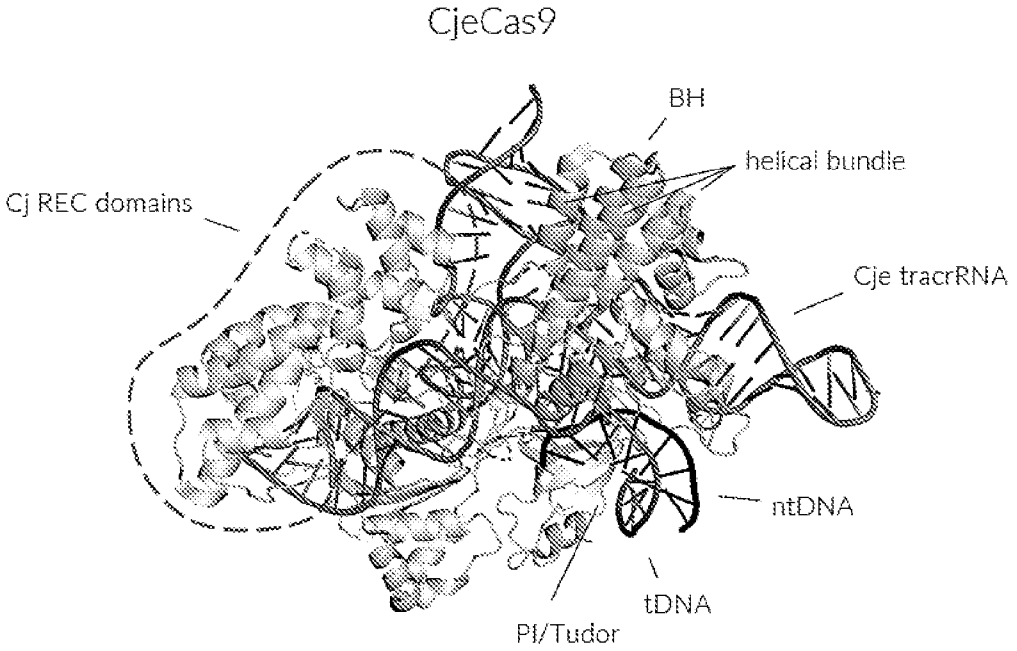

CjeCas9
BH
helical bundle
Cj REC domains
Cje tracrRNA
ntDNA
tDNA
PI/Tudor

FIG. 1A

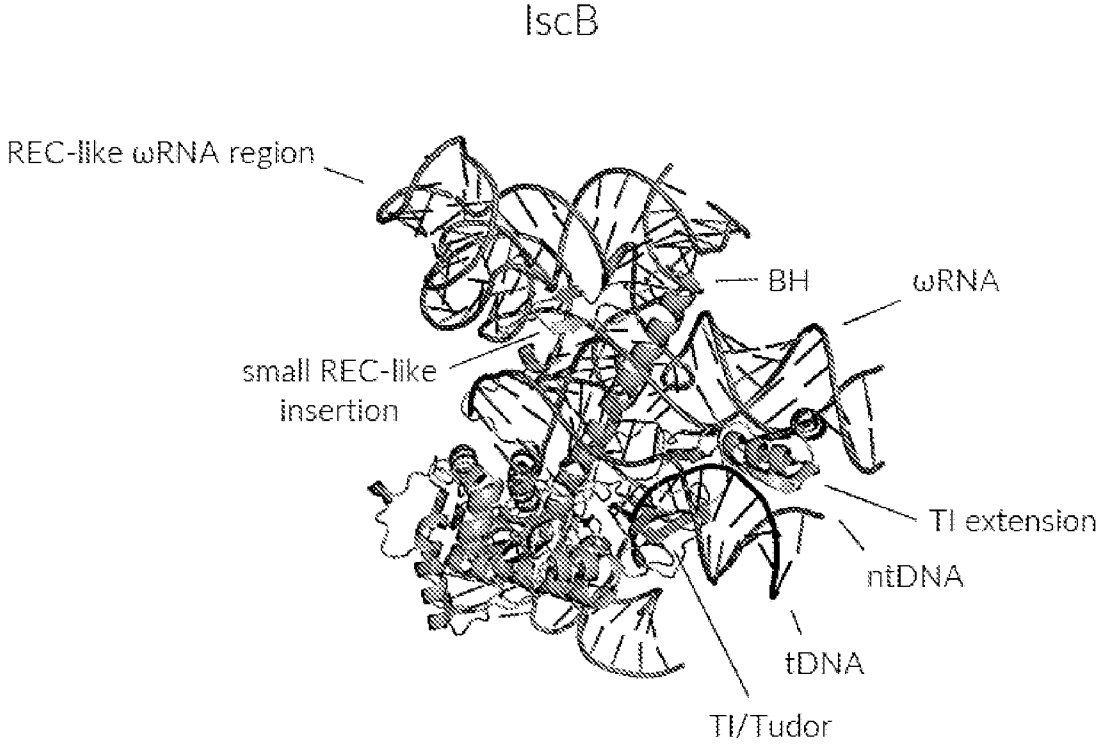

IscB
REC-like ωRNA region
BH
ωRNA
small REC-like insertion
TI extension
ntDNA
tDNA
TI/Tudor

FIG. 1B

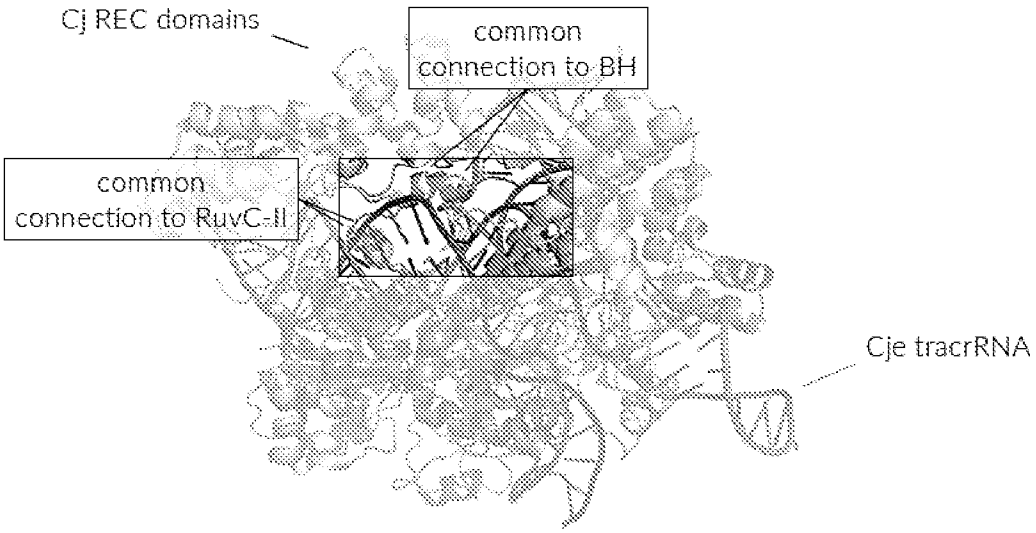

Cj REC domains
common connection to BH
common connection to RuvC-II
Cje tracrRNA

FIG. 2A

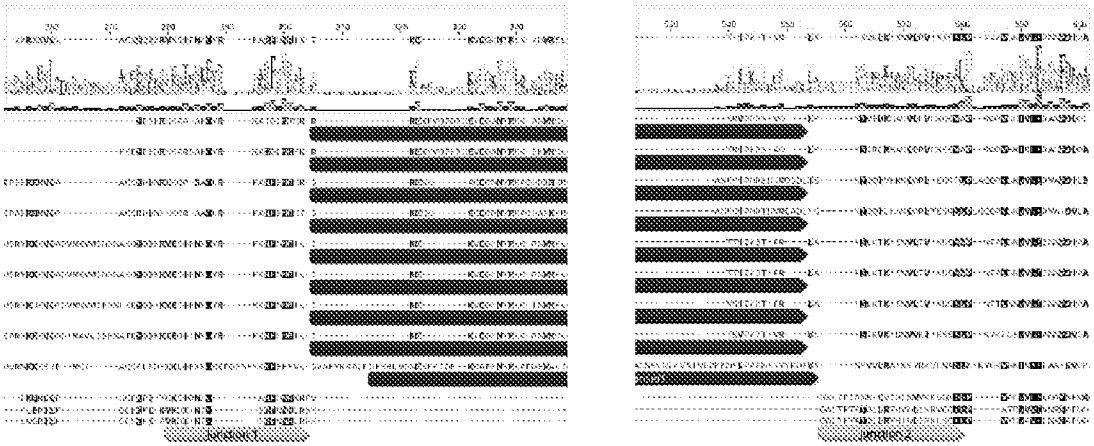


FIG. 2B

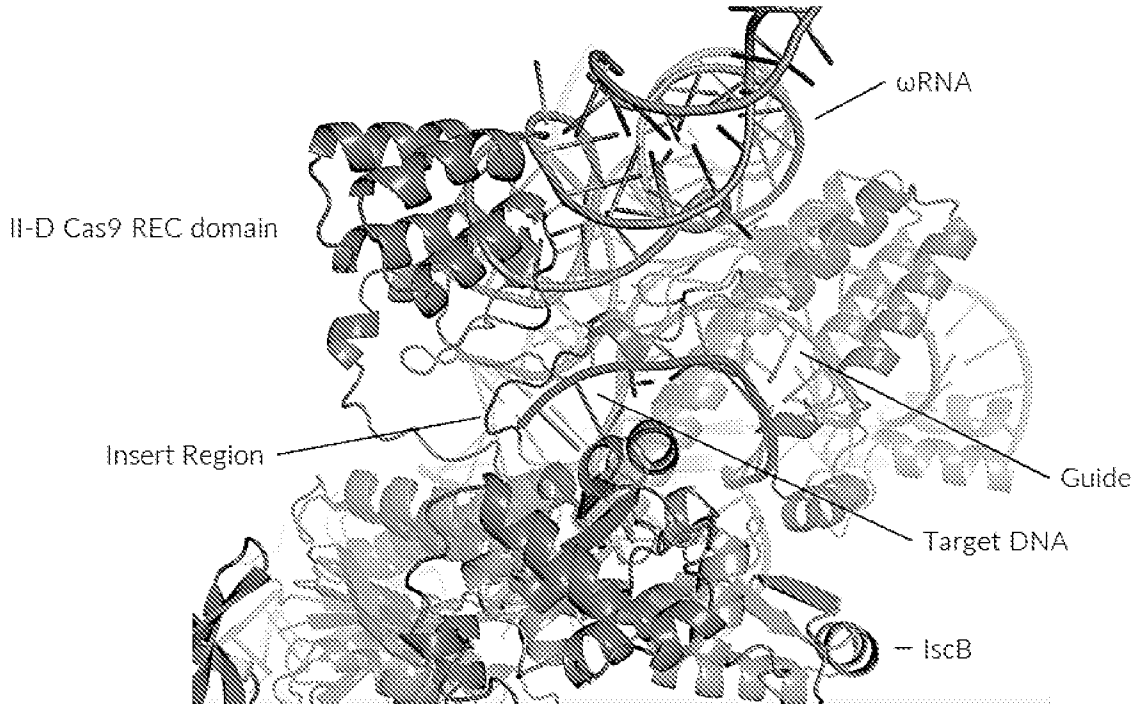

ωRNA
II-D Cas9 REC domain
Insert Region
Guide
Target DNA
IscB

FIG. 3

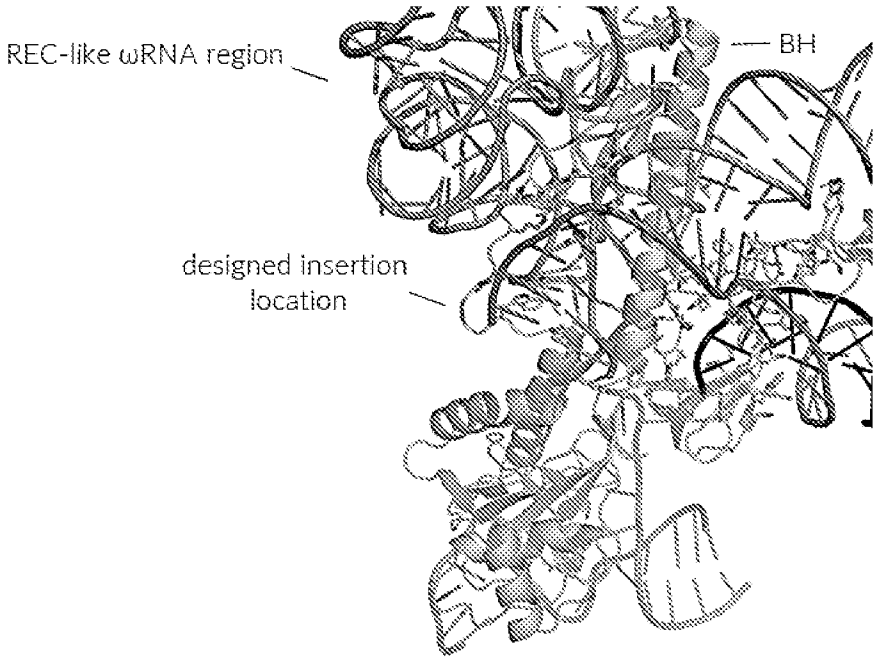

REC-like ωRNA region
BH
designed insertion location

FIG. 4A

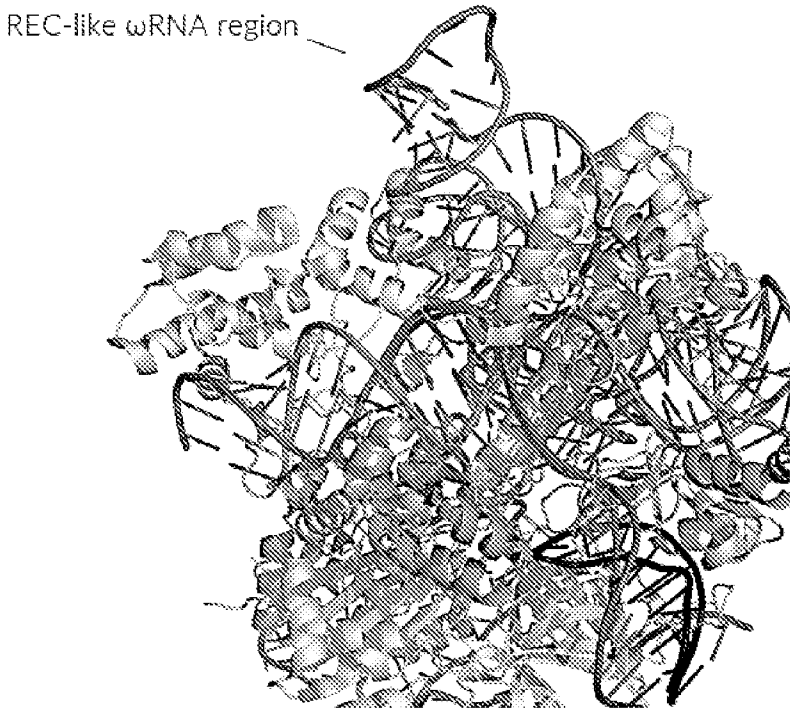

REC-like ωRNA region

FIG. 4B

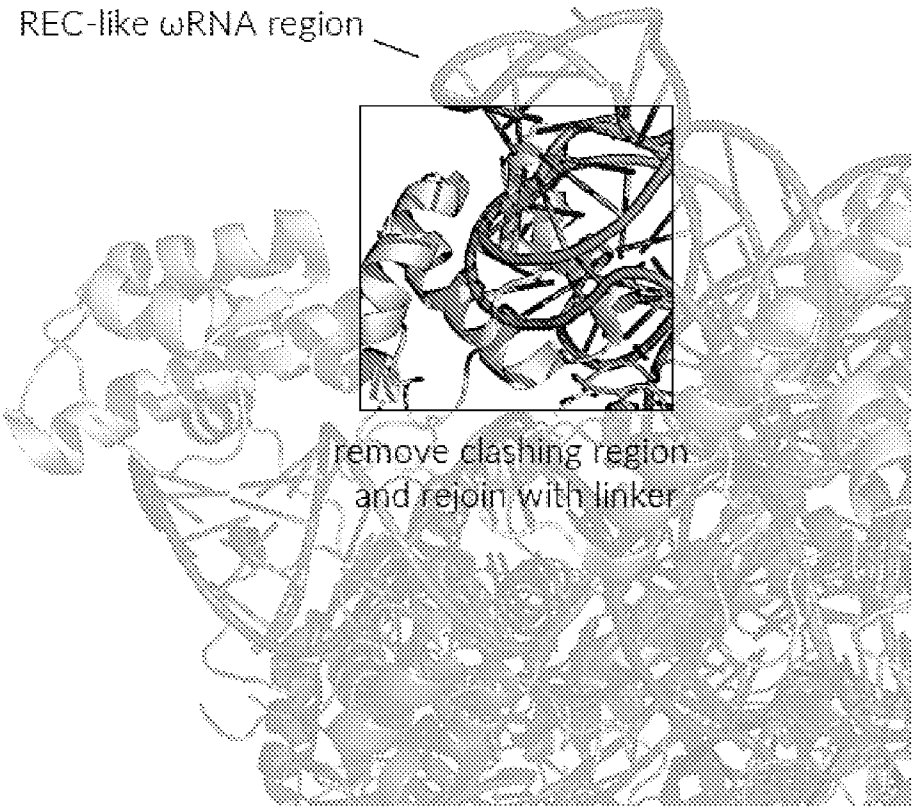

REC-like ωRNA region
remove clashing region
and rejoin with linker

FIG. 4C

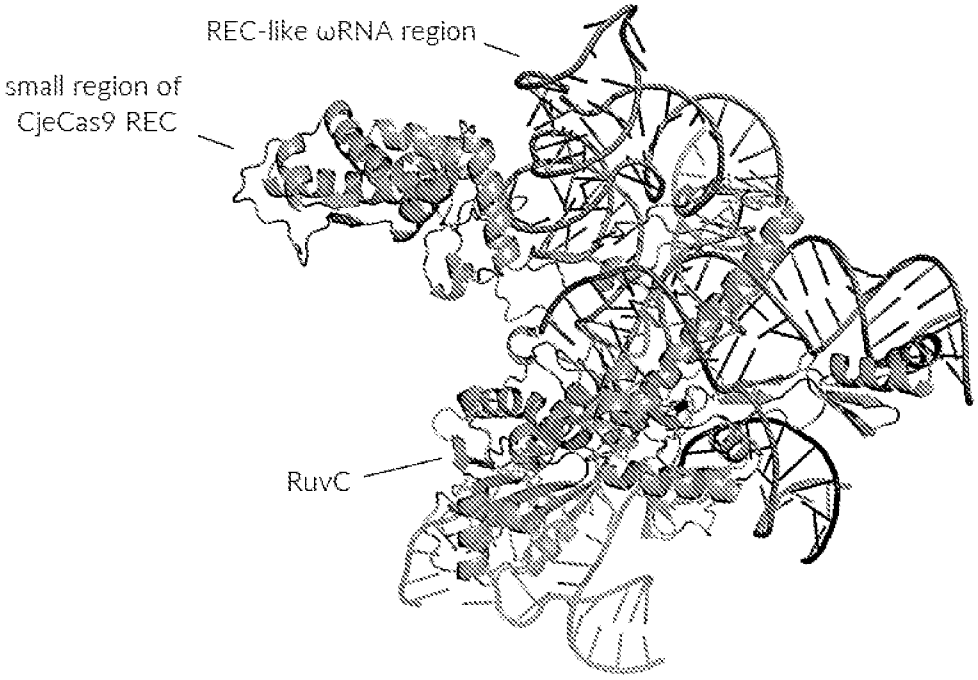

REC-like ωRNA region
small region of
CjeCas9 REC
RuvC

FIG. 5A

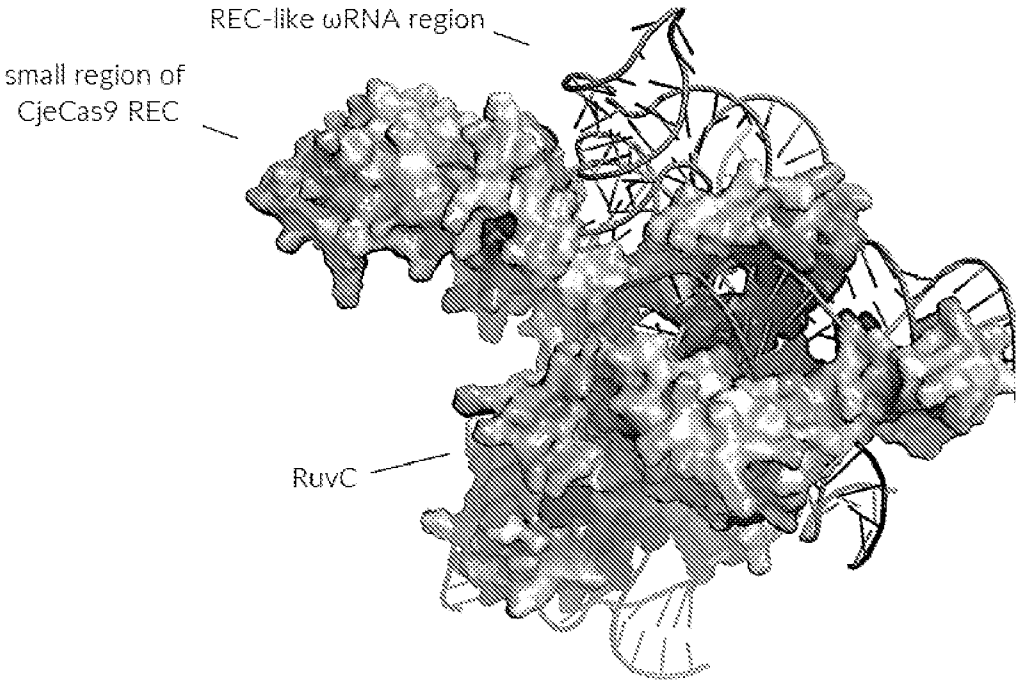

REC-like ωRNA region
small region of
CjeCas9 REC
RuvC

FIG. 5B

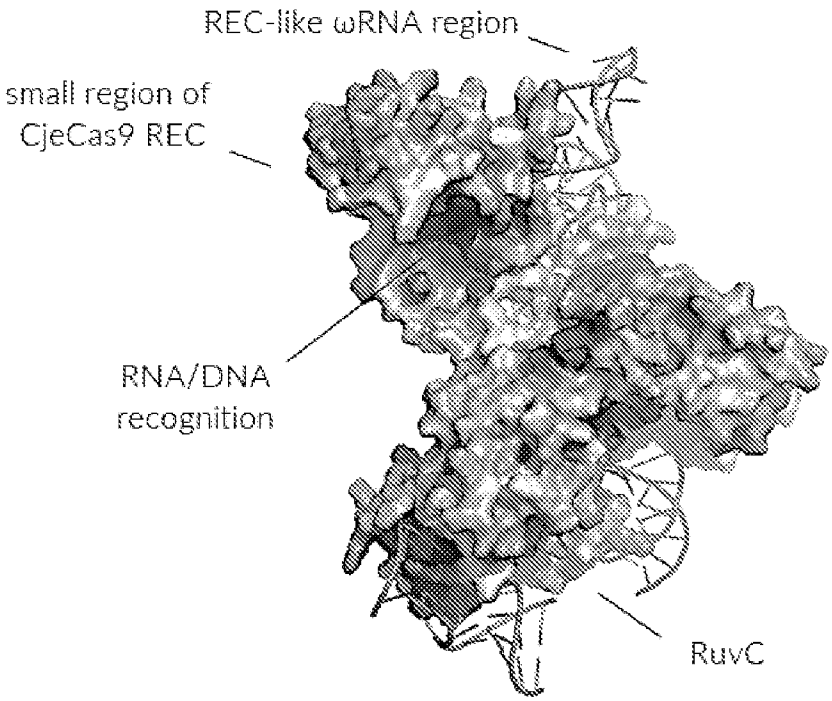

REC-like ωRNA region
small region of
CjeCas9 REC
RNA/DNA
recognition
RuvC

FIG. 6


RNase H

FIG. 7

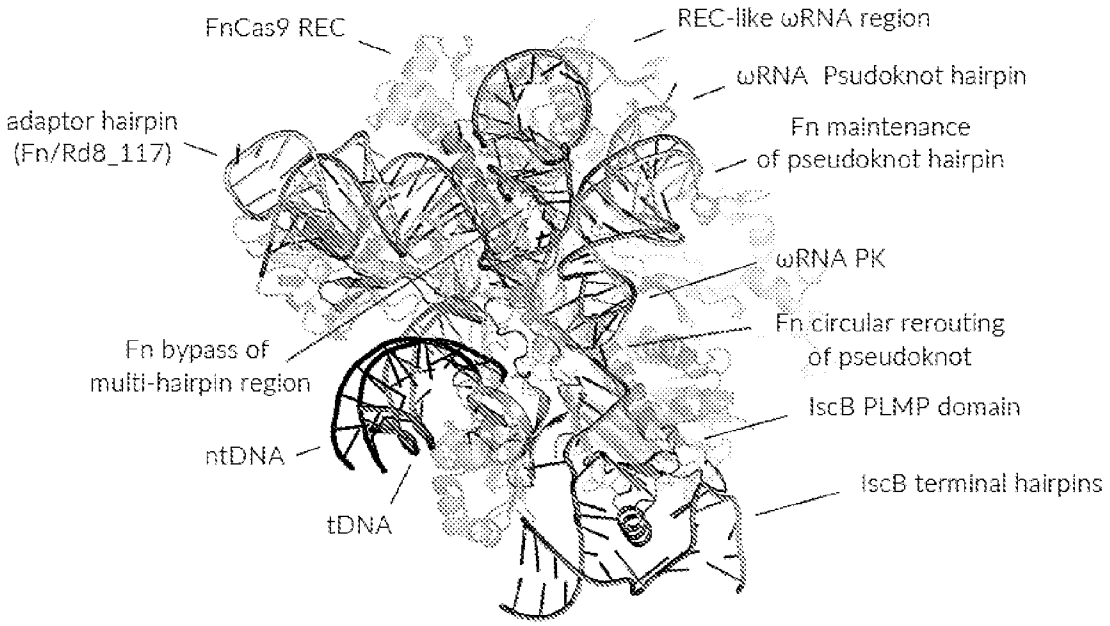

FnCas9 REC
REC-like ωRNA region
ωRNA Psudoknot hairpin
adaptor hairpin
(Fn/Rd8_117)
Fn maintenance
of pseudoknot hairpin
ωRNA PK
Fn circular rerouting
of pseudoknot
Fn bypass of
multi-hairpin region
IscB PLMP domain
ntDNA
IscB terminal hairpins
tDNA

FIG. 8

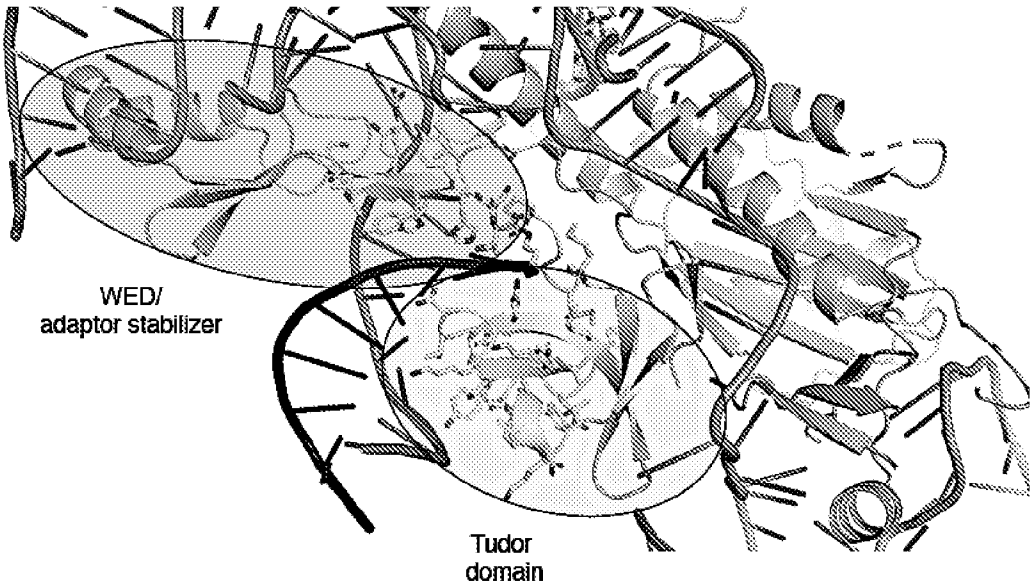

WED/
adaptor stabilizer
Tudor
domain

FIG. 9

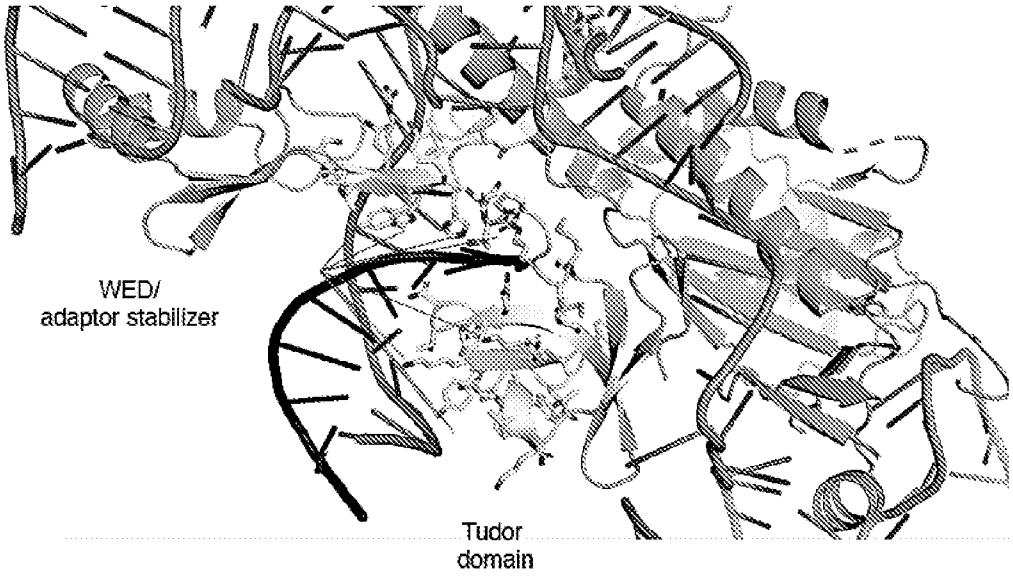

WED/
adaptor stabilizer
Tudor
domain

FIG. 10

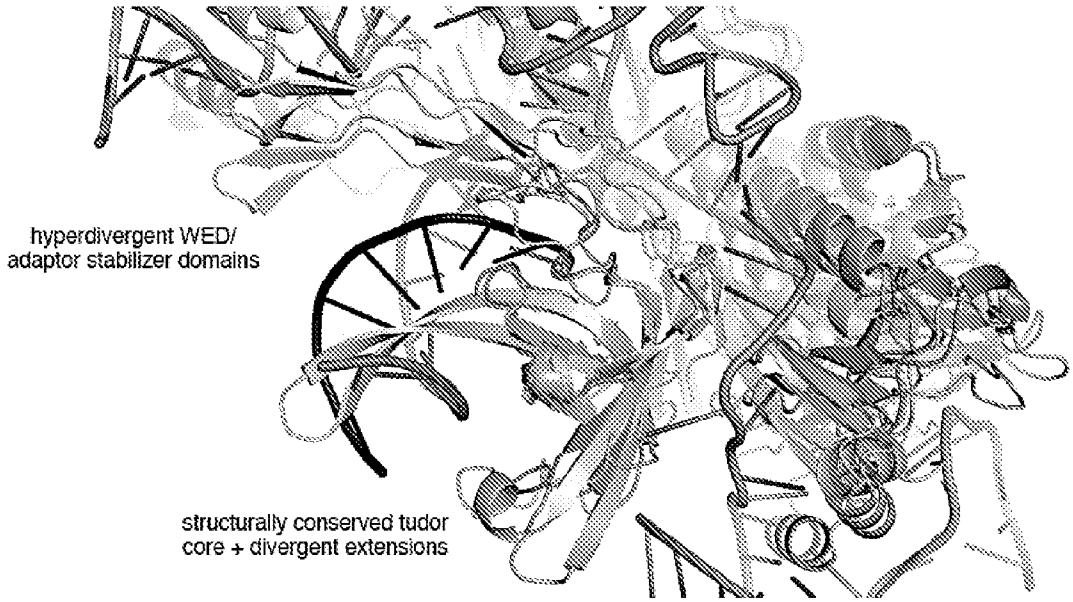

hyperdivergent WED/
adaptor stabilizer domains
structurally conserved tudor
core + divergent extensions

FIG. 11

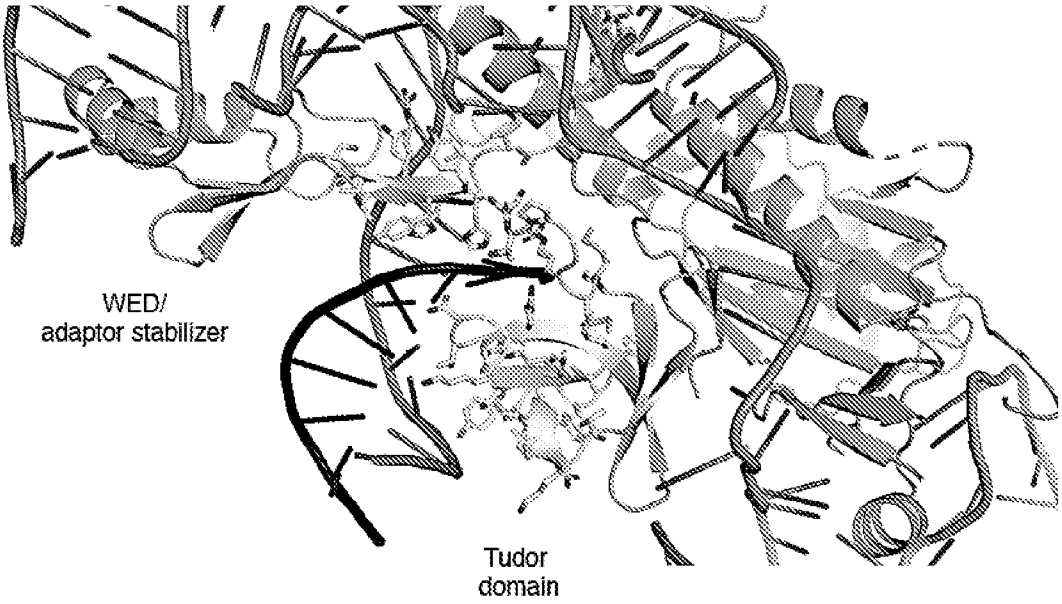


FIG. 12

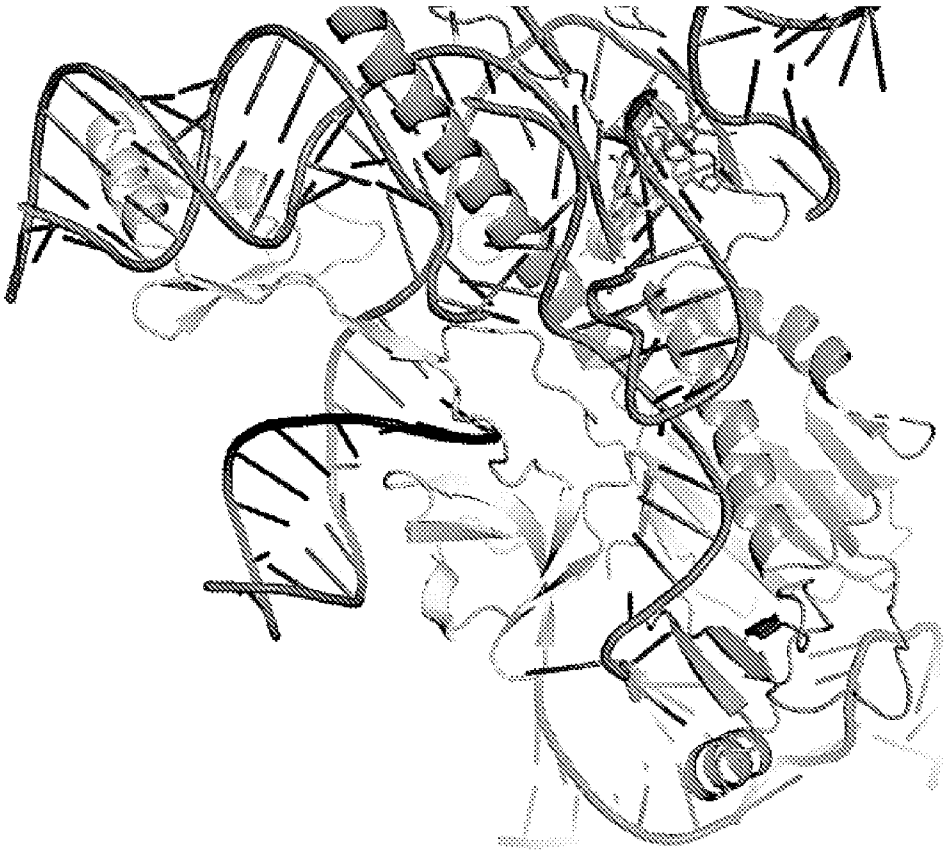


FIG. 13A

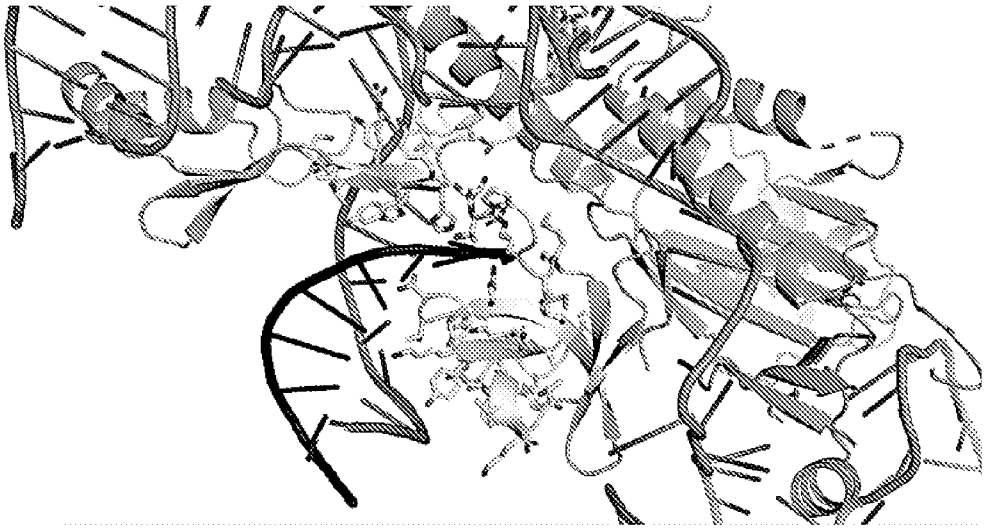


FIG. 13B

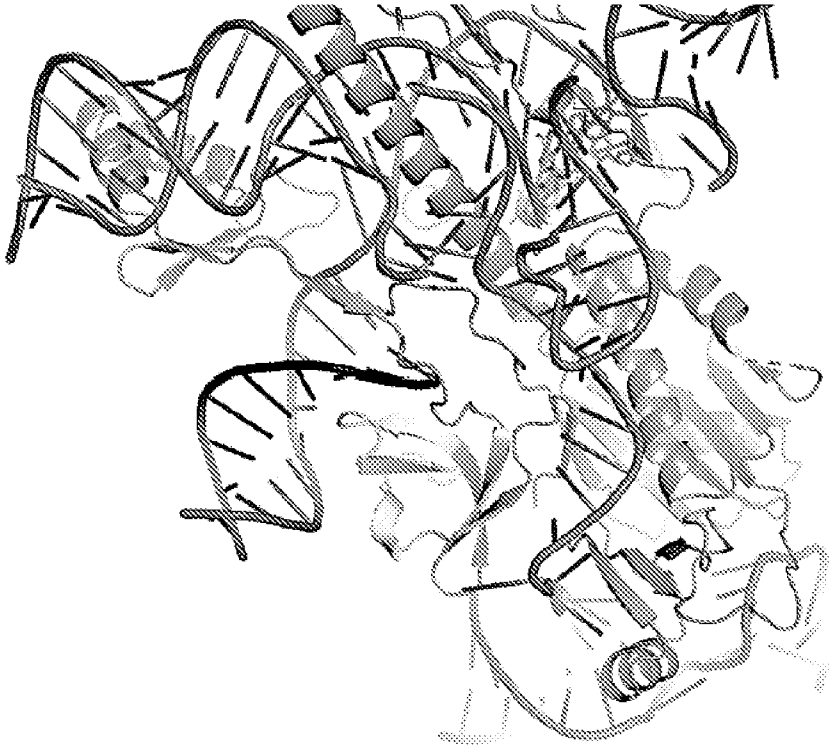


FIG. 13C

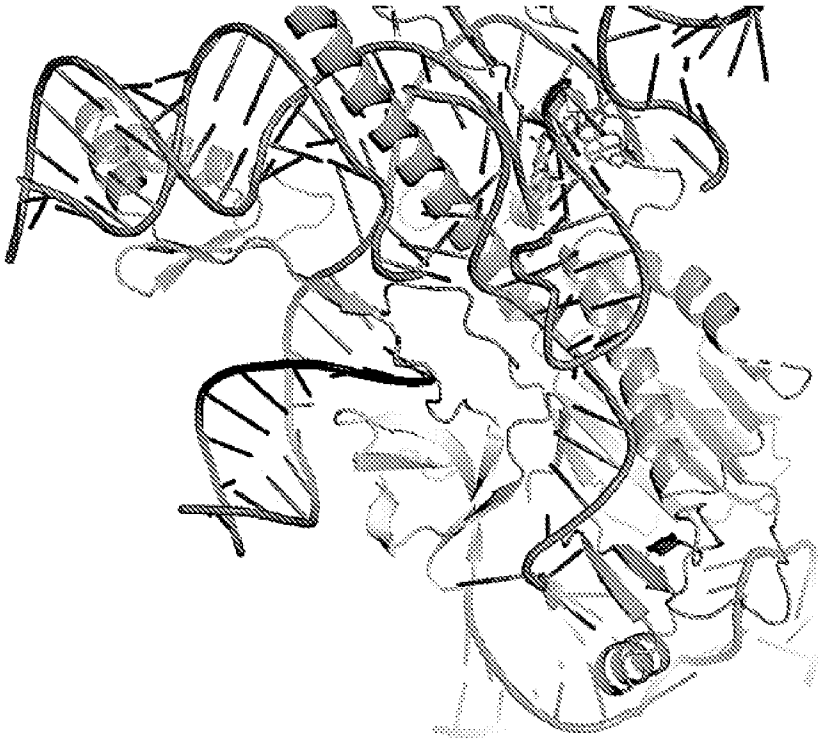


FIG. 13D

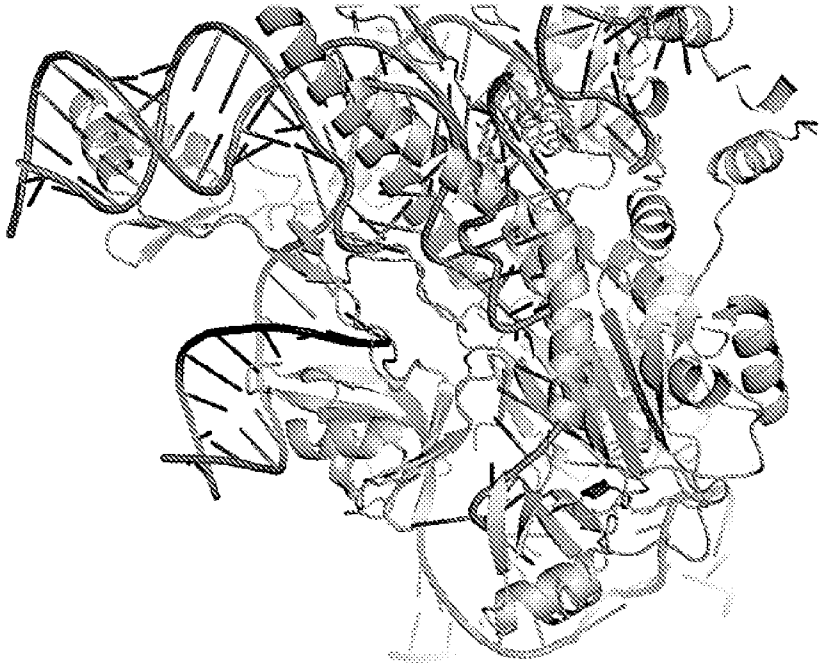


FIG. 13E

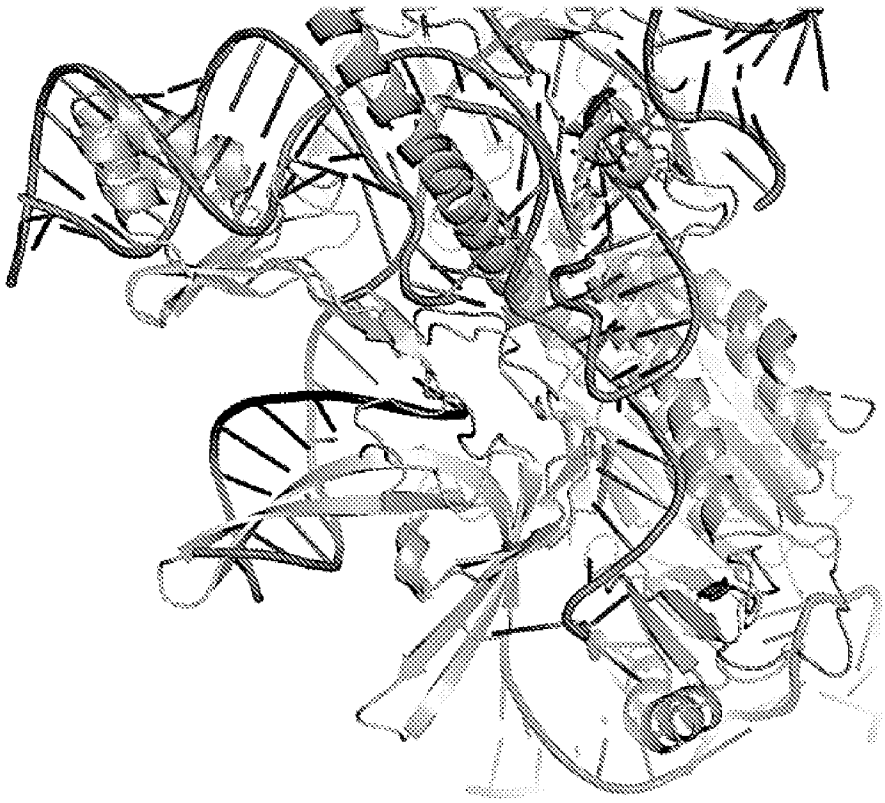


FIG. 13F

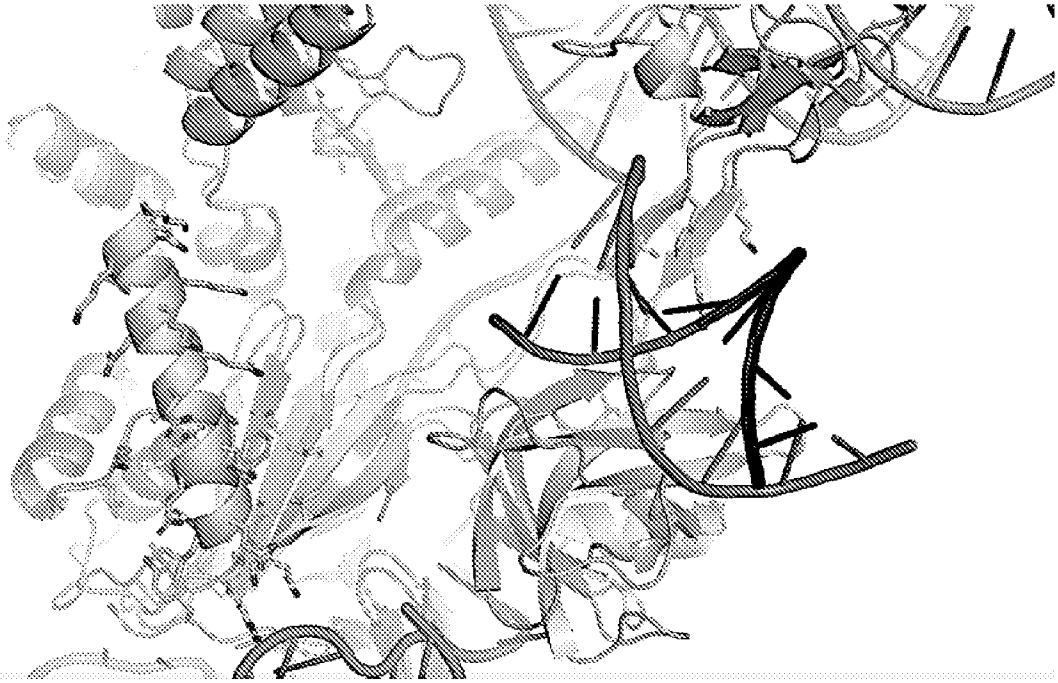


FIG. 14



FIG. 15

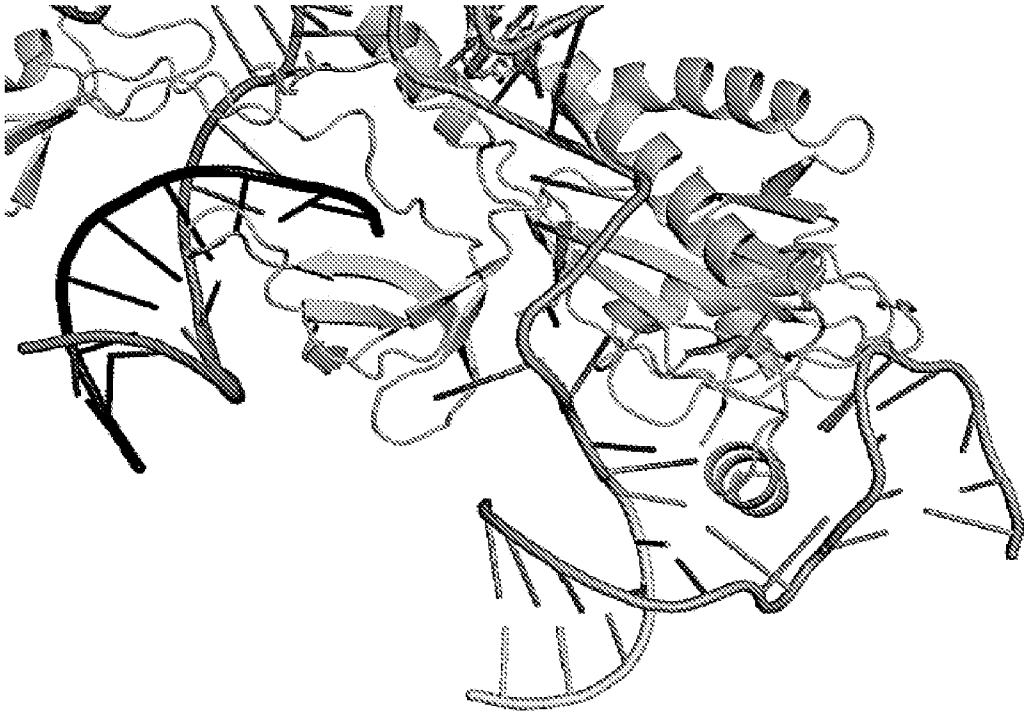


FIG. 16

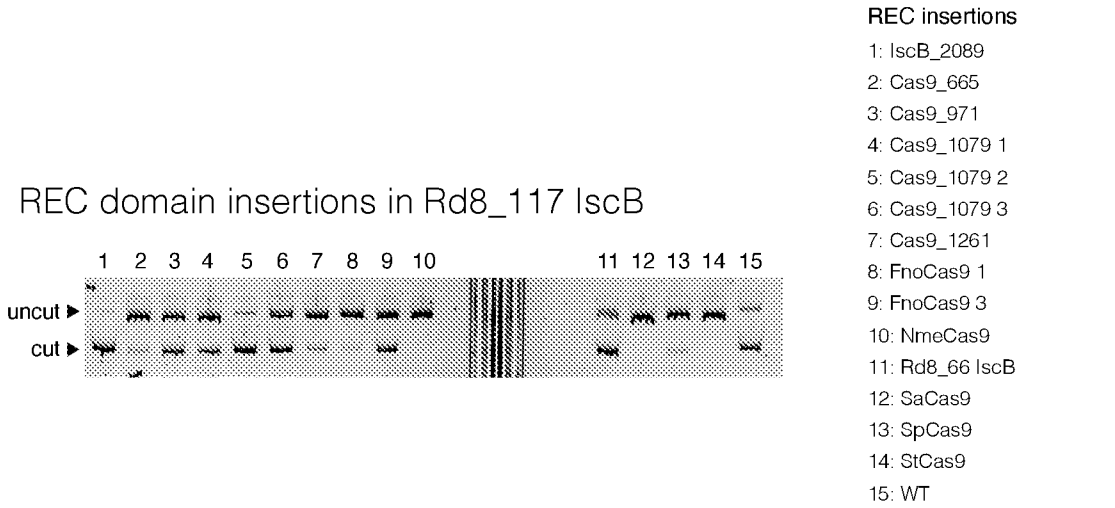

REC domain insertions in Rd8_117 IscB
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
uncut
cut
REC insertions
1: IscB_2089
2: Cas9_665
3: Cas9_971
4: Cas9_1079 1
5: Cas9_1079 2
6: Cas9_1079 3
7: Cas9_1261
8: FnoCas9 1
9: FnoCas9 3
10: NmeCas9
11: Rd8_66 IscB
12: SaCas9
13: SpCas9
14: StCas9
15: WT

FIG. 17

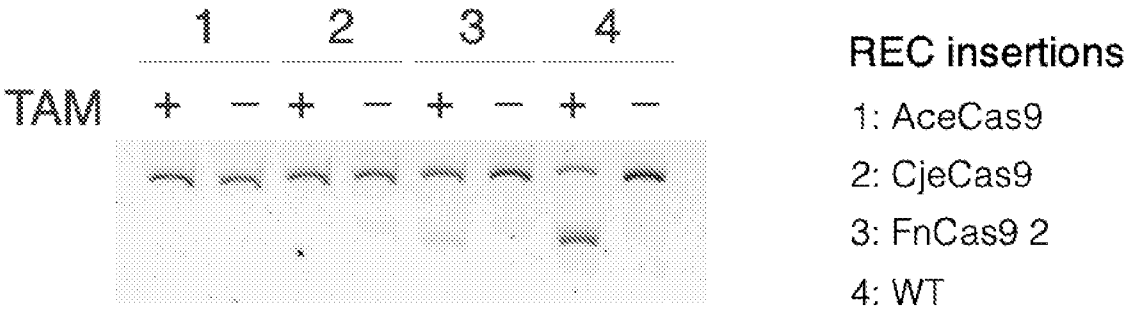

1 2 3 4
TAM + − + − + − + −
REC insertions
1: AceCas9
2: CjeCas9
3: FnCas9 2
4: WT

FIG. 18

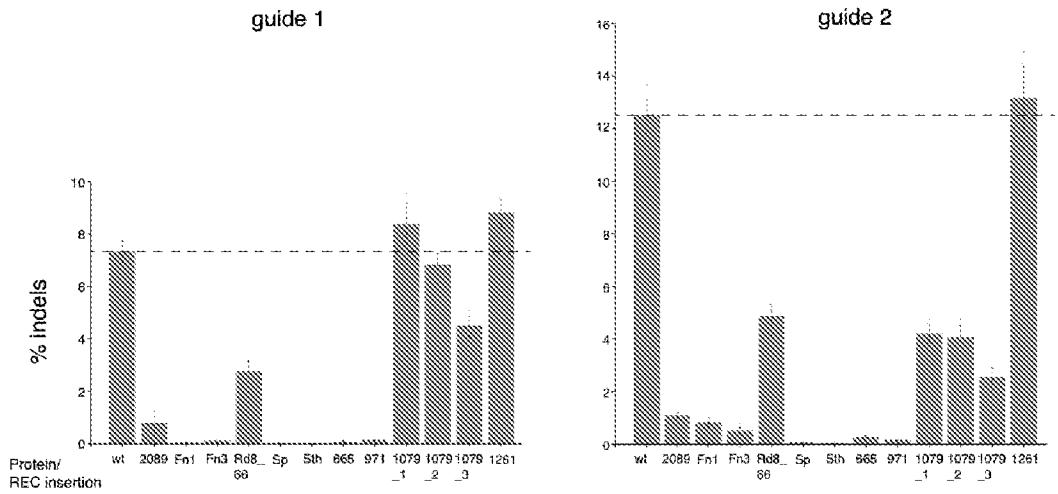


FIG. 19

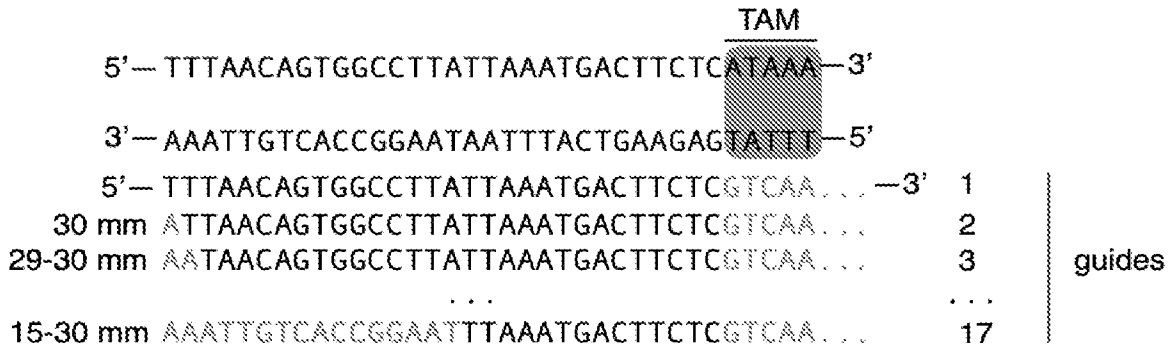


FIG. 20A

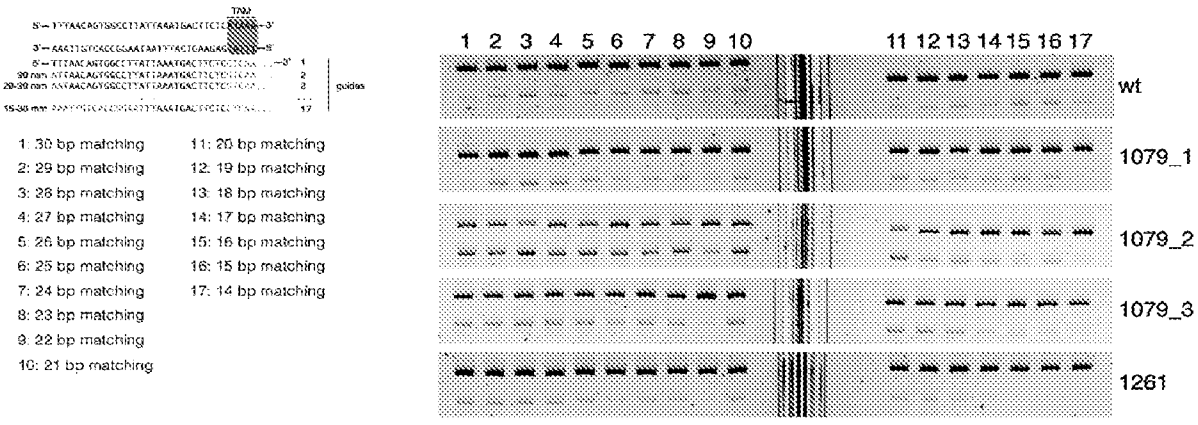


FIG. 20B

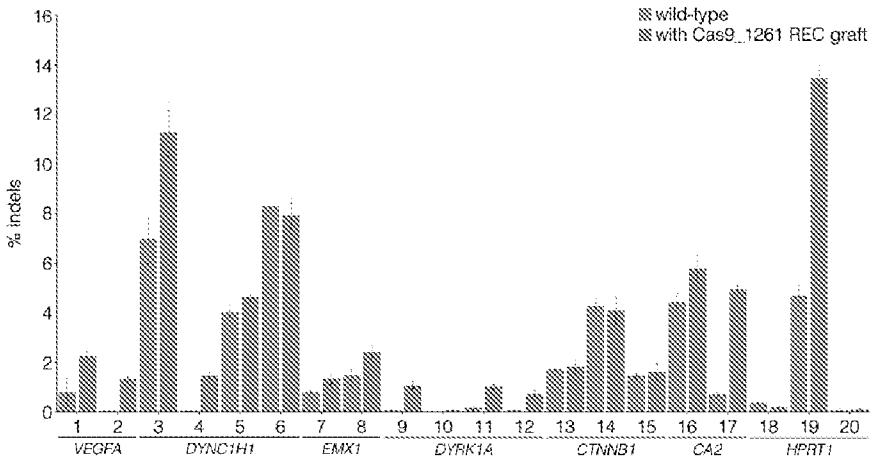


FIG. 21

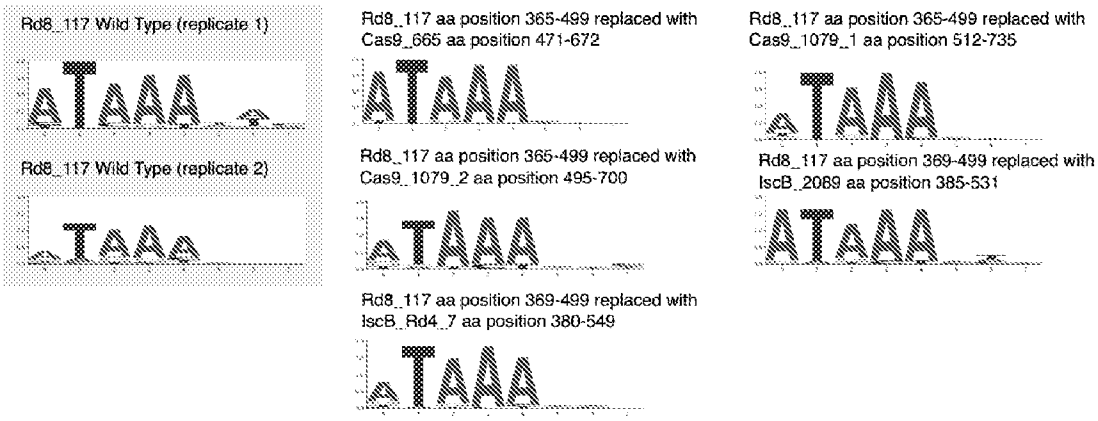


FIG. 22A

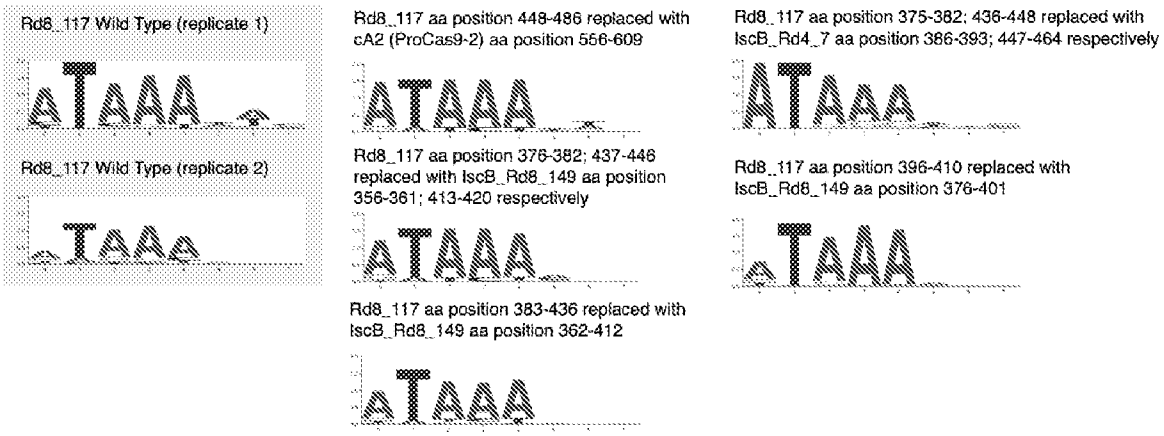


FIG. 22B

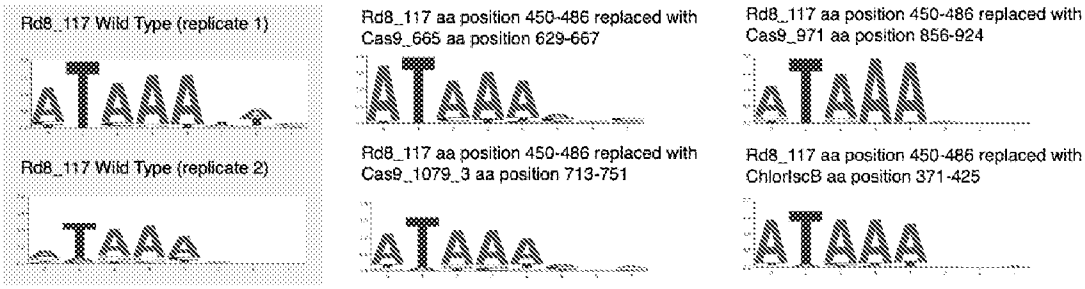


FIG. 22C

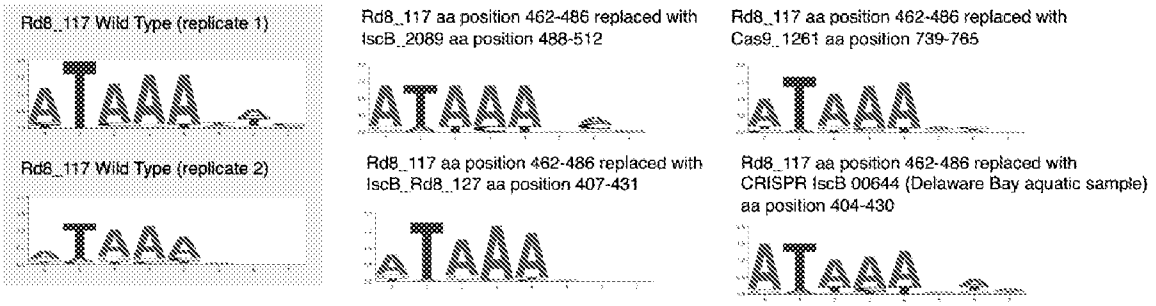


FIG. 22D

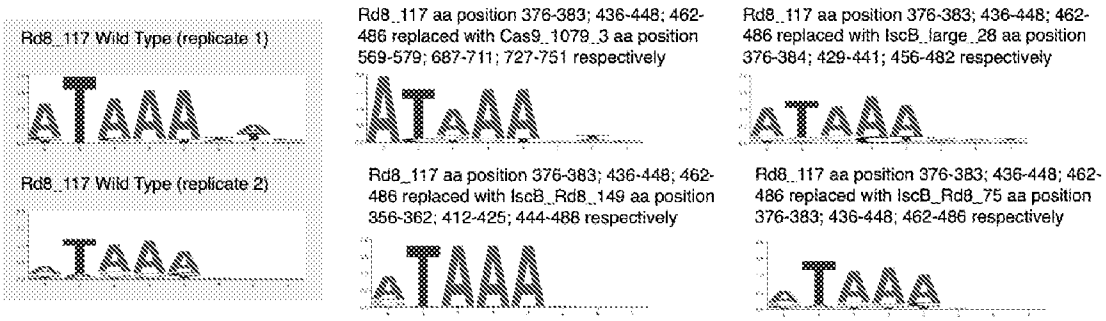


FIG. 22E

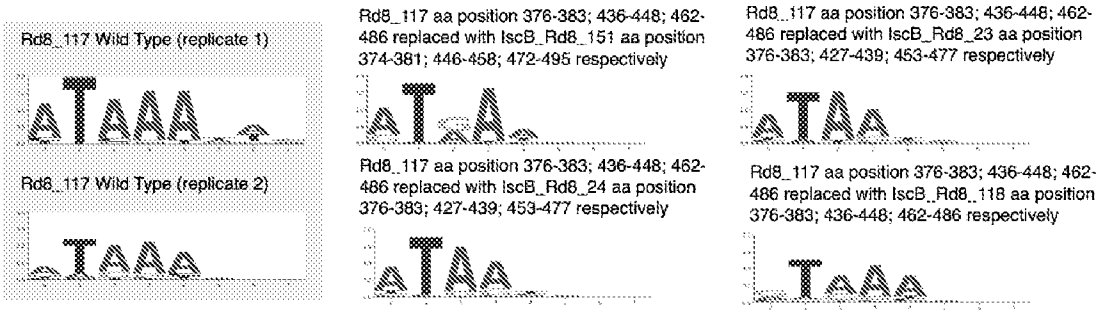


FIG. 22F

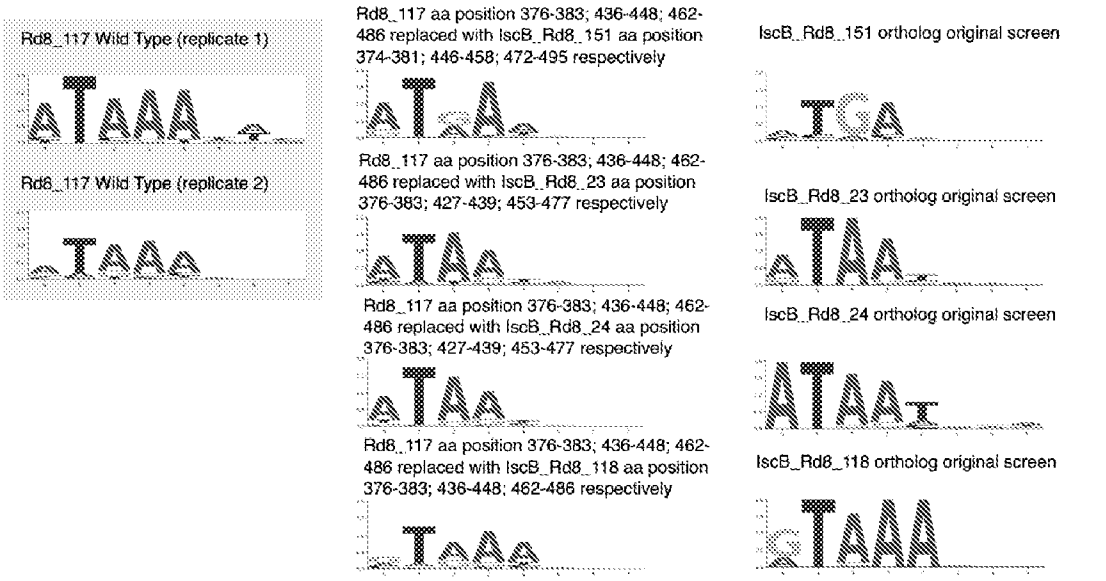


FIG. 23

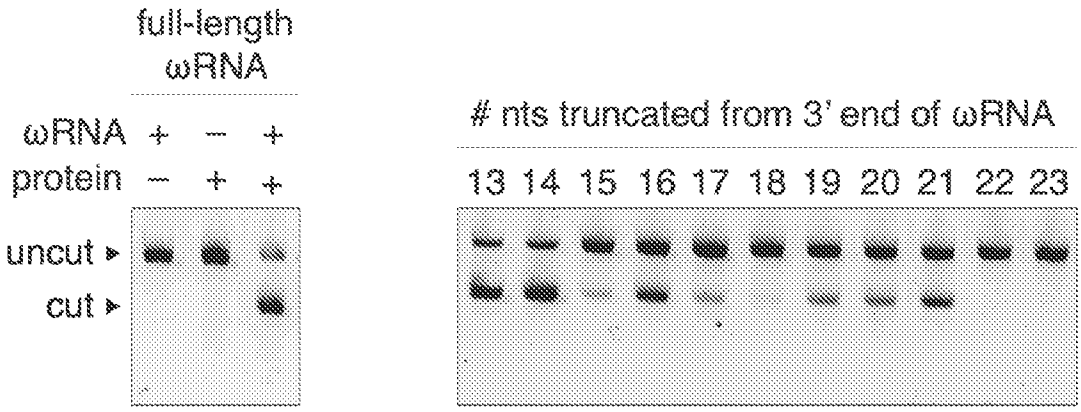
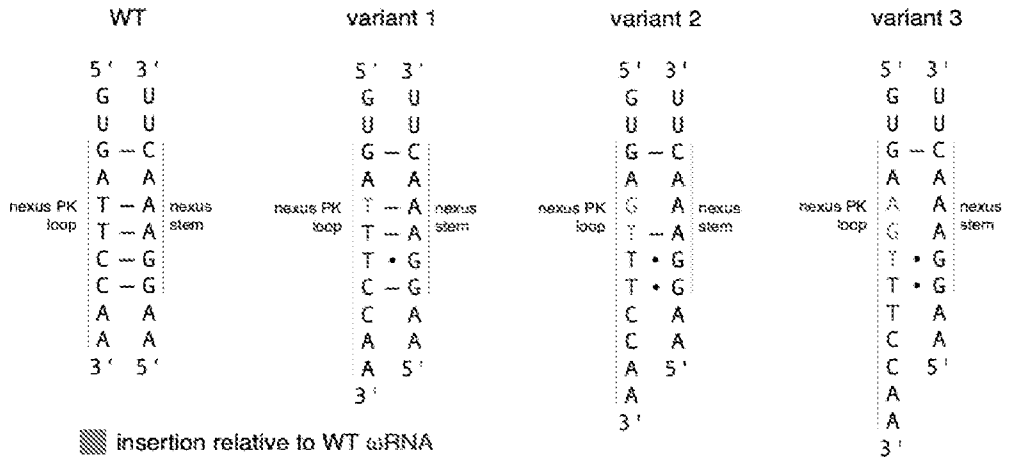


FIG. 24A



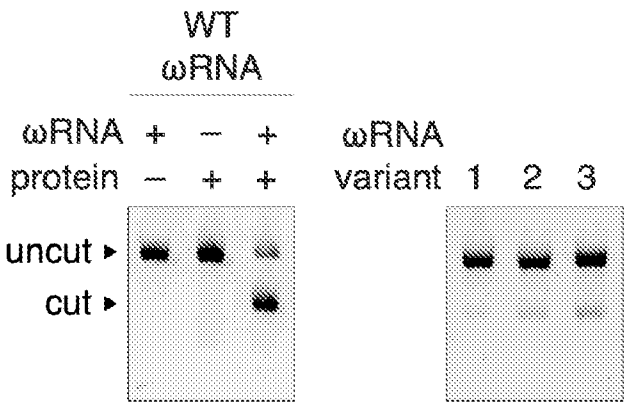
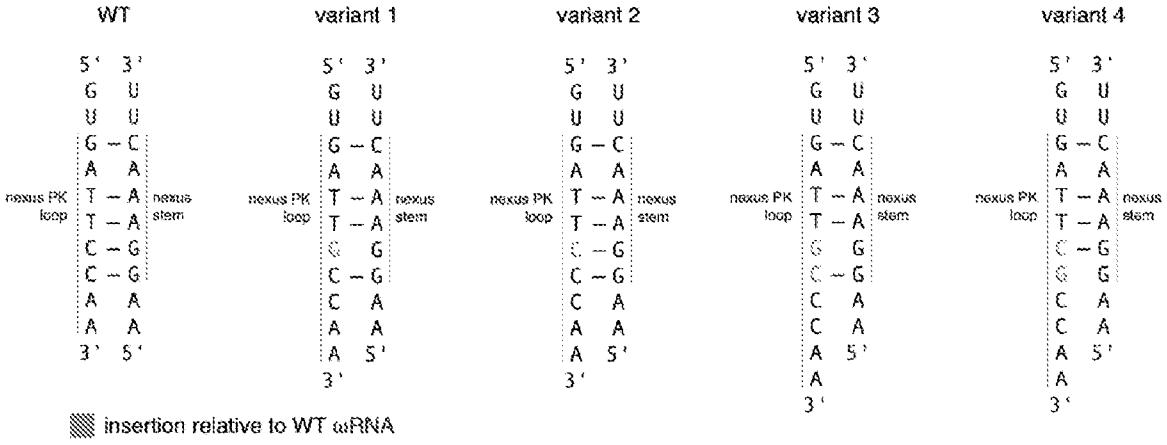


FIG. 24B


WT
variant 1
variant 2
variant 3
variant 4
nexus PK loop
nexus stem
insertion relative to WT ωRNA

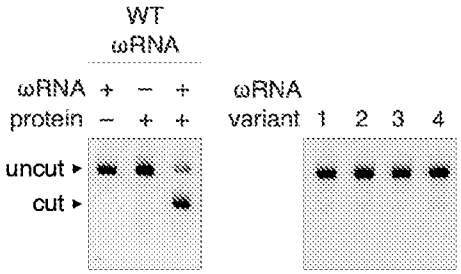
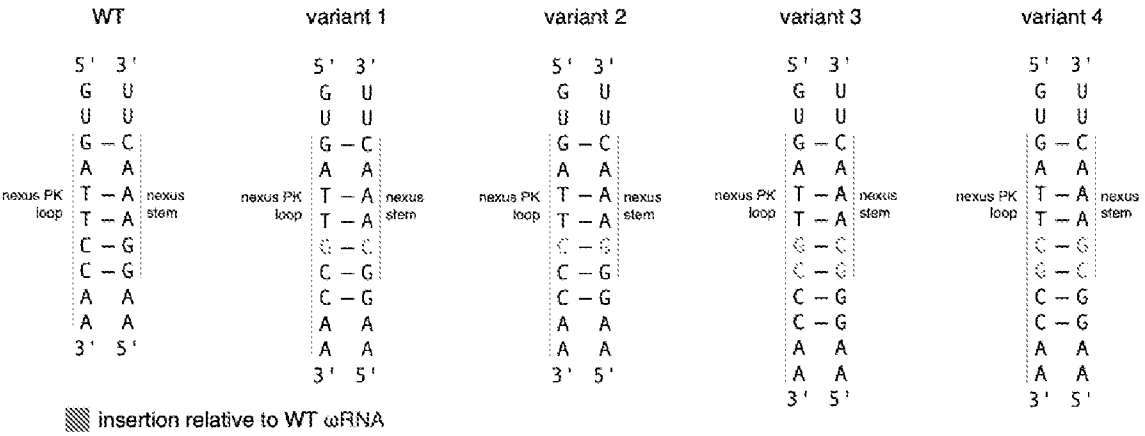

WT ωRNA
ωRNA + − +
protein − + +
uncut
cut
ωRNA variant 1 2 3 4

FIG. 24C


WT
variant 1
variant 2
variant 3
variant 4
nexus PK loop
nexus stem
insertion relative to WT ωRNA

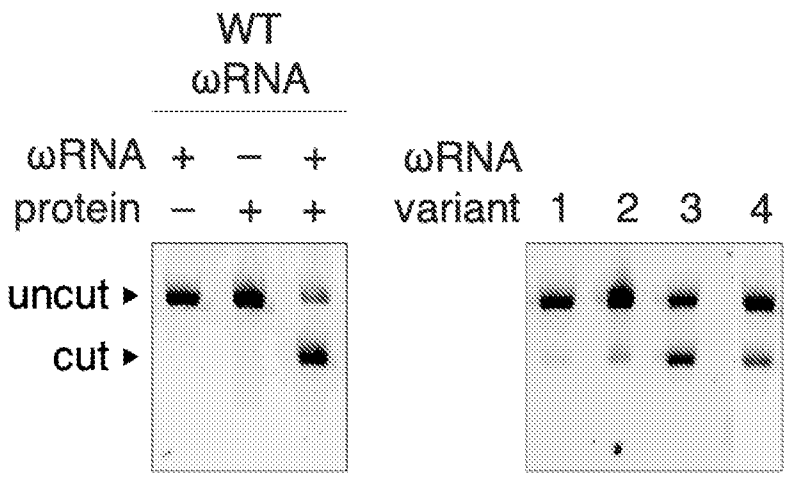

WT
ωRNA
ωRNA + − +
protein − + +
ωRNA
variant 1 2 3 4
uncut
cut

FIG. 24D

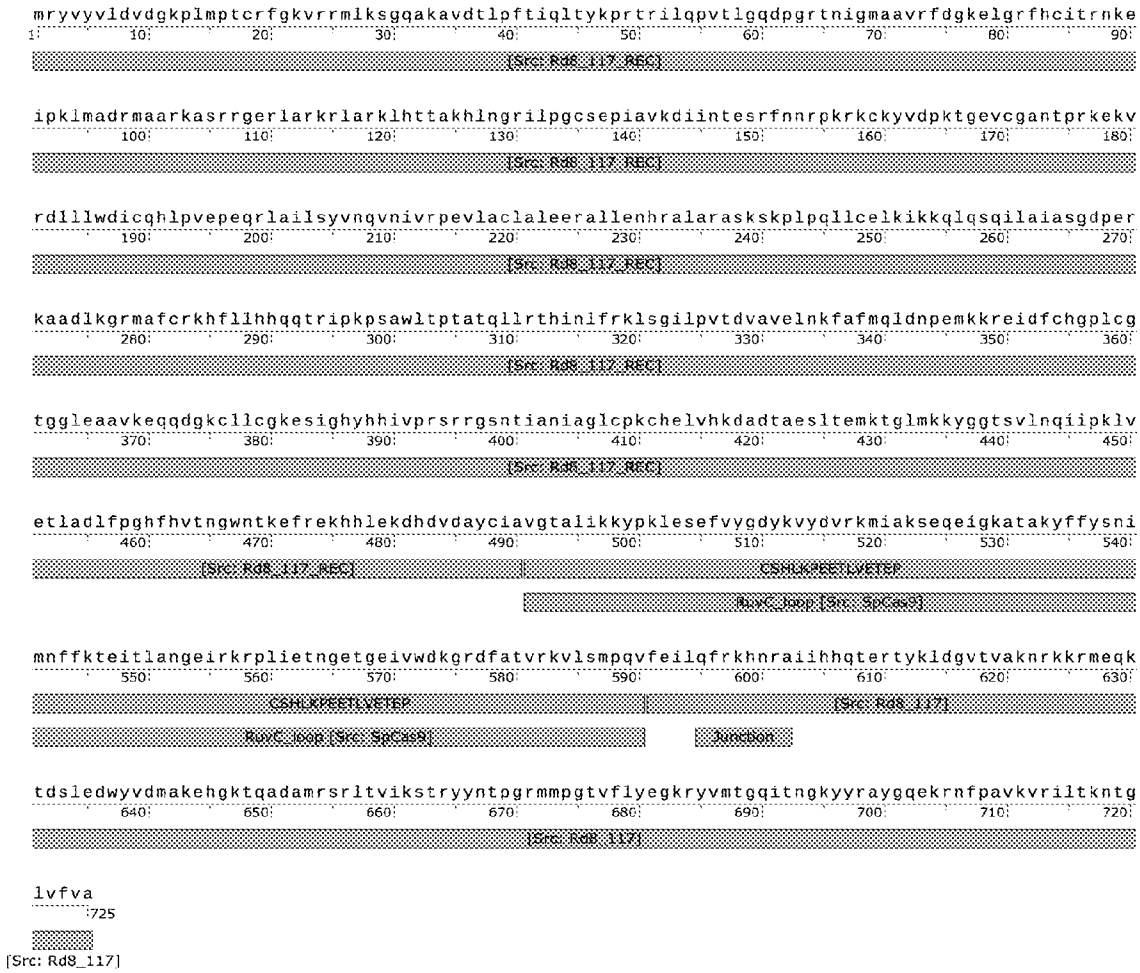

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnke
1 10 20 30 40 50 60 70 80 90
[Src: Rd8_117_REC]
ipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekv
100 110 120 130 140 150 160 170 180
[Src: Rd8_117_REC]
rdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqllcelkikkqlqsqilaiasgdper
190 200 210 220 230 240 250 260 270
[Src: Rd8_117_REC]
kaadlkgrmafcrkhfllhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcg
260 290 300 310 320 330 340 350 360
[Src: Rd8_117_REC]
tggleaavkeqqdgkcllcgkesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklv
370 380 390 400 410 420 430 440 450
[Src: Rd8_117_REC]
etladlfpghfhvtngwntkefrekhhlekdhdvdayciavgtalikkypklesefvygdykvydvrkmiakseqeigkatakyffysni
460 470 480 490 500 510 520 530 540
[Src: Rd8_117_REC]
CSHLKPEETLVETEP
RuvC_loop [Src: SpCas9]
mnffkteitlangeirkrplietngetgeivwdkgrdfatvrkvlsmpqvfeilqfrkhnraiihhqtertyklдgvtvaknrkkrmeqk
550 560 570 580 590 600 610 620 630
CSHLKPEETLVETEP
[Src: Rd8_117]
RuvC_loop [Src: SpCas9]
Junction
tdsledwyvdmakehgktqadamrsrltvikstryyntpgrmmpgtvflyegkryvmtgqitngkyyraygqekrnfpavkvriltkntg
640 650 660 670 680 690 700 710 720
[Src: Rd8_117]
lvfva
725
[Src: Rd8_117]

FIG. 25

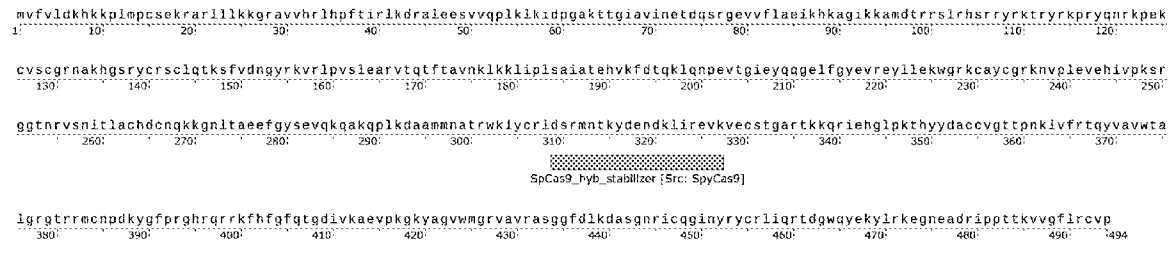

mvfvldkhkkplmpcsekrarlllkkgravvhrlhpftirlkdraleesvvqplklkidpgakttgiavinetdqsrgevvflaeikhkagikkamdtrrslrhsrryrktryrkpryqnrkpek
1 10 20 30 40 50 60 70 80 90 100 110 120
cvscgrnakhgsrycrsclqtksfvdngyrkvrlpvslearvtqtftavnklkkliplsaiatehvkfdtqklqnpevtgieyqqgelfgyevreyllekwgrkcaycgrknvplevehivpksr
130 140 150 160 170 180 190 200 210 220 230 240 250
ggtnrvsnitlachdcnqkkgnltaeefgysevqkqakqplkdaammnatrwkiycridsrmntkydendklirevkvecstgartkkqriehglpkthyydaccvgttpnkivfrtqyvavwta
260 270 280 290 300 310 320 330 340 350 360 370
SpCas9_hyb_stabilizer [Src: SpyCas9]
lgrgtrrmcnpdkygfprghrgrrkfhfgfqtgdivkaevpkgkyagvwmgrvavrasggfdlkdasgnricqginyrycrliqrtdgwqvekylrkegneadripptkvvgflrcvp
380 390 400 410 420 430 440 450 460 470 480 490 494

FIG. 26



FIG. 27

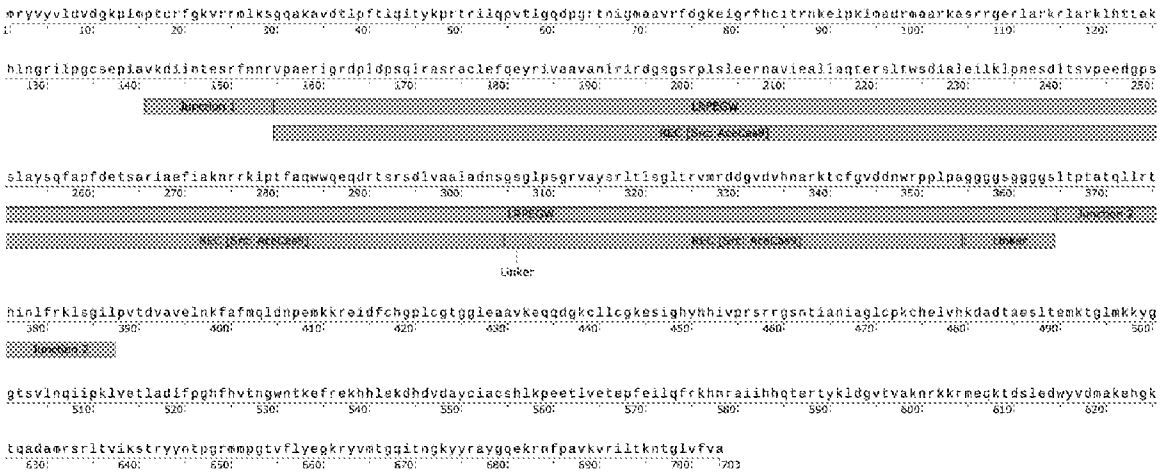


FIG. 28



FIG. 29



FIG. 30



FIG. 31



FIG. 32



FIG. 33



FIG. 34



FIG. 35



FIG. 36

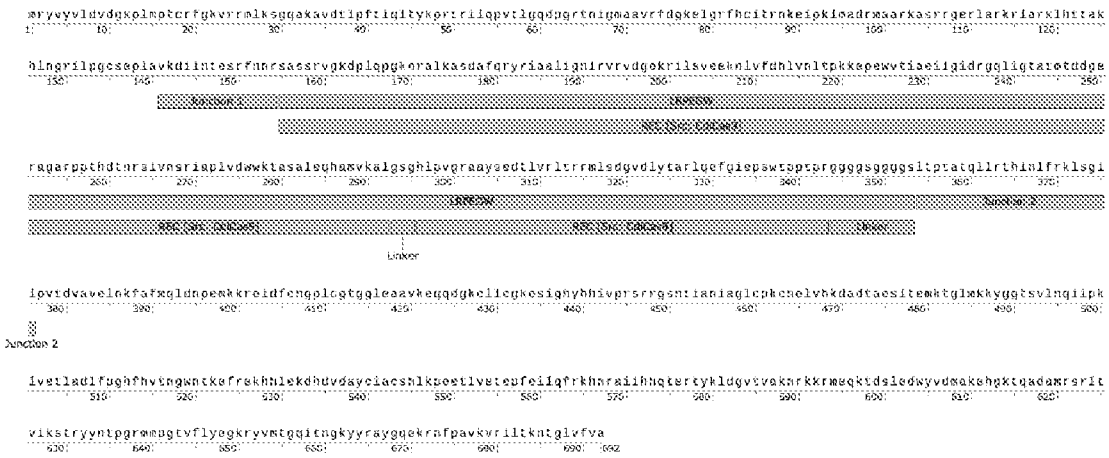


FIG. 37



FIG. 38



FIG. 39

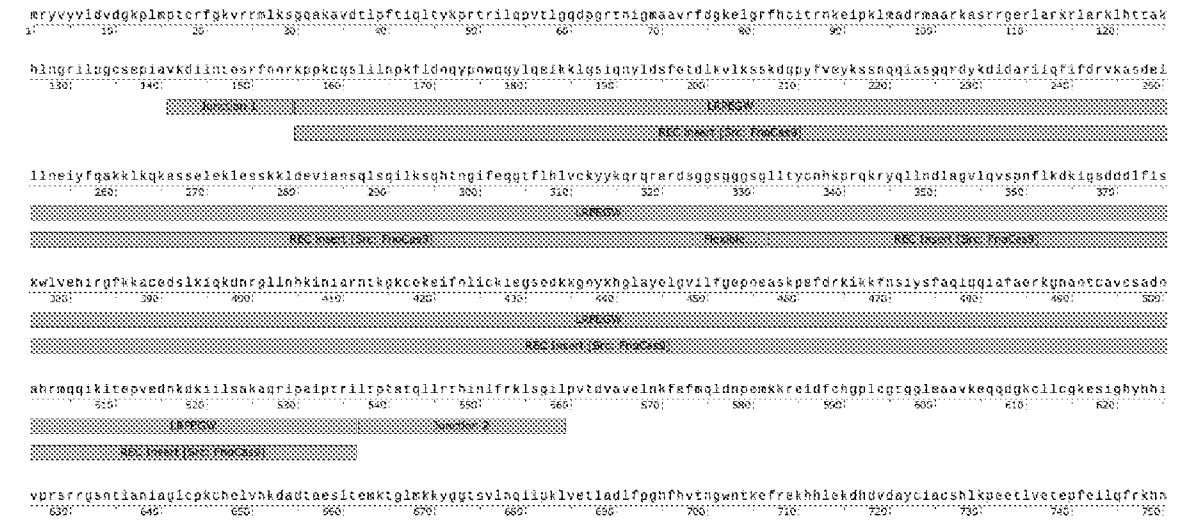


FIG. 40



FIG. 41



FIG. 42



FIG.43

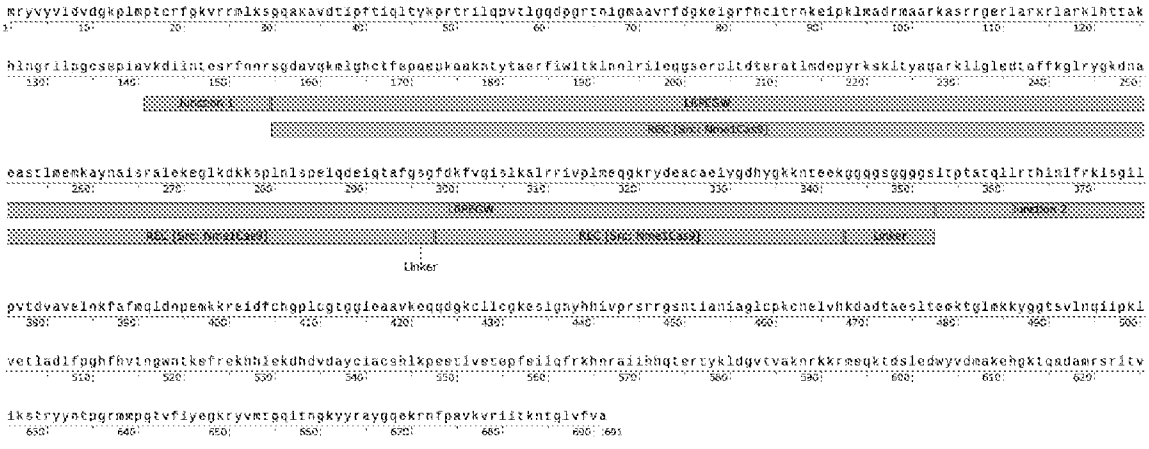


FIG. 44



FIG. 45



FIG. 46

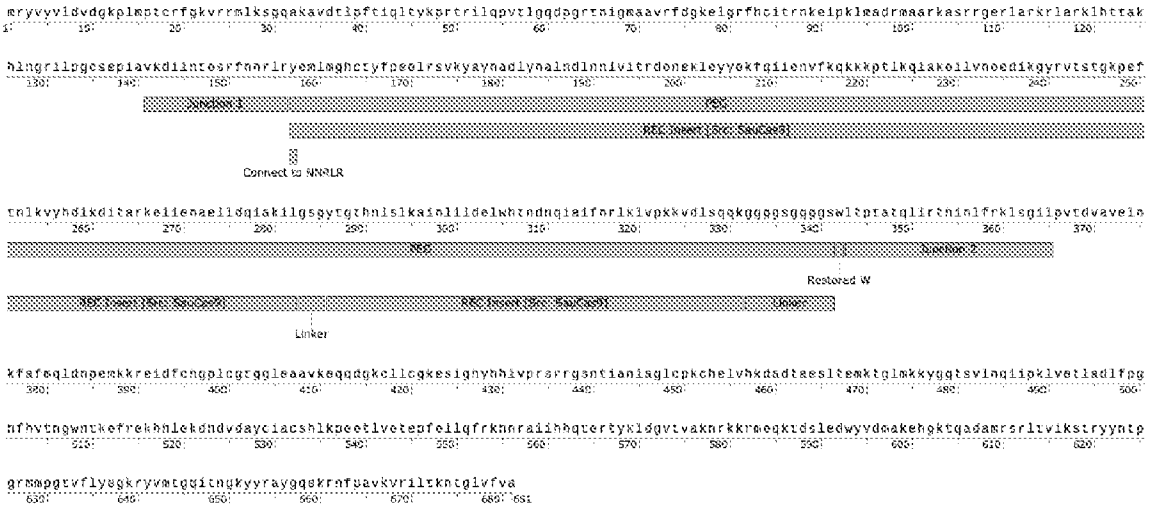


FIG. 47

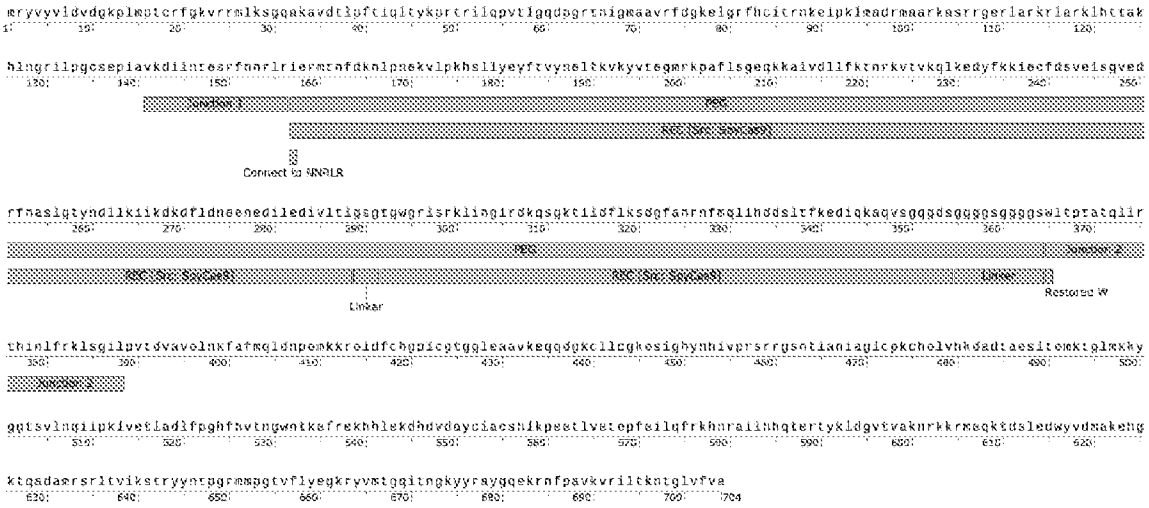

Connect to RNP/LR
Linker
Restored W

FIG. 48


SpCas9_hyb_stabilizer (Src: SpyCas9)

FIG. 49



FIG. 50

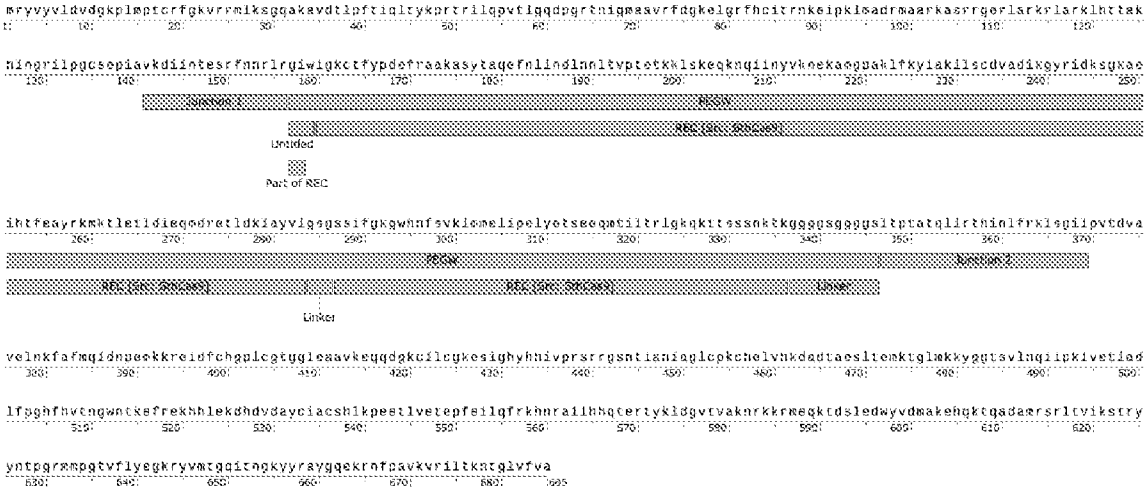


FIG. 51

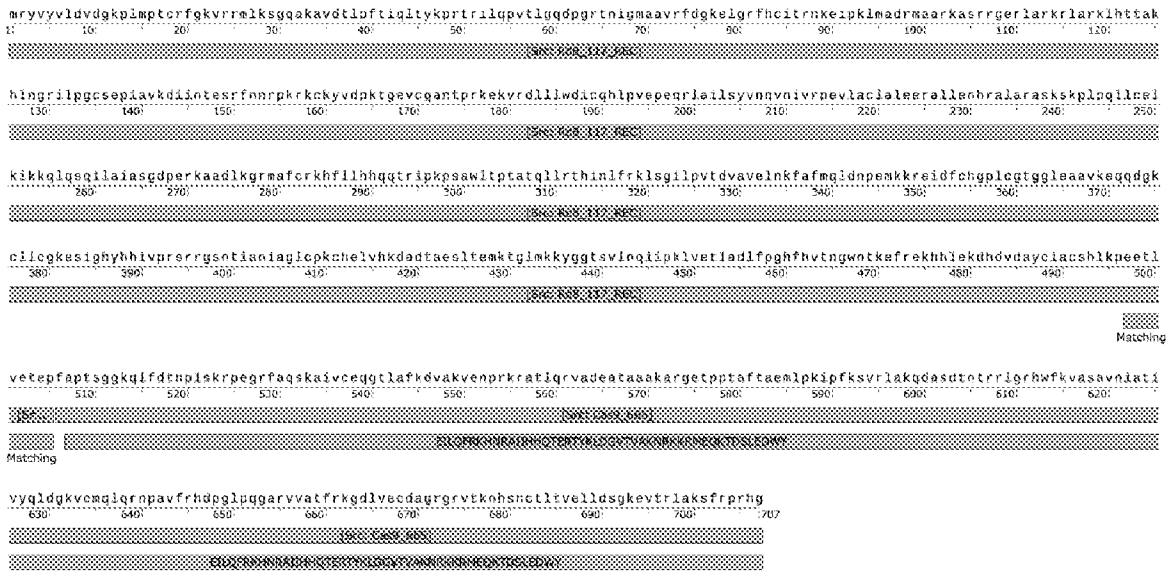

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftigltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttak
hlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdlllwdicqhlpvepeqrlailsyvnqvnivrnevlaclaleeralleehraiaraskskplpqllcei
kikkqlqsqilaiasgdperkaadlkgrmaferkhfilhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
cilcgkesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesltemktgimkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetl
vetepfaptsgqkqlfdtnpiskrpegrfaqskaivceqgtlafkdvakvenprkratiqrvadeataaakargetpptaftaemlpkipfksvrlakqdasdtotrrlqrhwfkvasavniati
vyqldgkvcmqlqrnpavfrhdpglpqgarvvatfrkgdlvecdagrgrvtkohsnctltvelldsgkevtrlaksfrprhg

FIG. 52

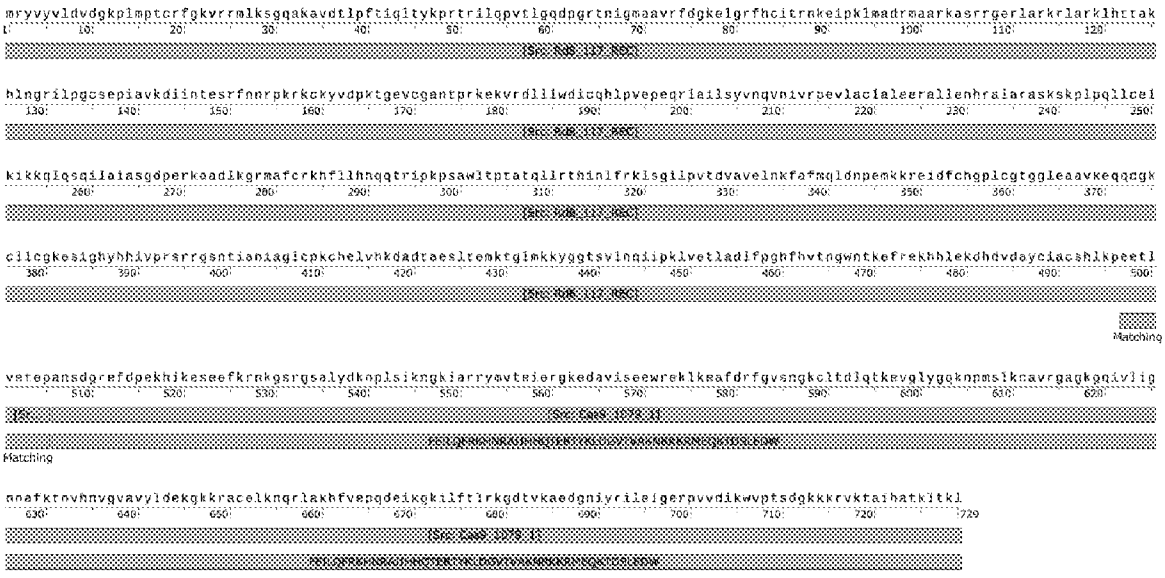

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftigltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttak
hlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdlliwdicqhlpvepeqrlailsyvnqvnivrnevlaclaleeralleehraiaraskskplpqllcei
kikkqlqsqilaiasgdperkaadlkgrmaferkhfilhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
cilcgkesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesltemktgimkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetl
vetepansdgrefdpekhikeseefkrnkgsrgsalydkqplsikngkiarryavteiergkedaviseewreklkeafdrfgvsngkcltdlqtkevglygeknpmslknavrgagkqnivlig
anafktnvhnvgvavyldekgkkracelkngrlakhfvepgdeikgkilftlrkgdtvkaedgniyrileigerpvvdikwvptsdgkkkrvktaihatkltkl

FIG. 53

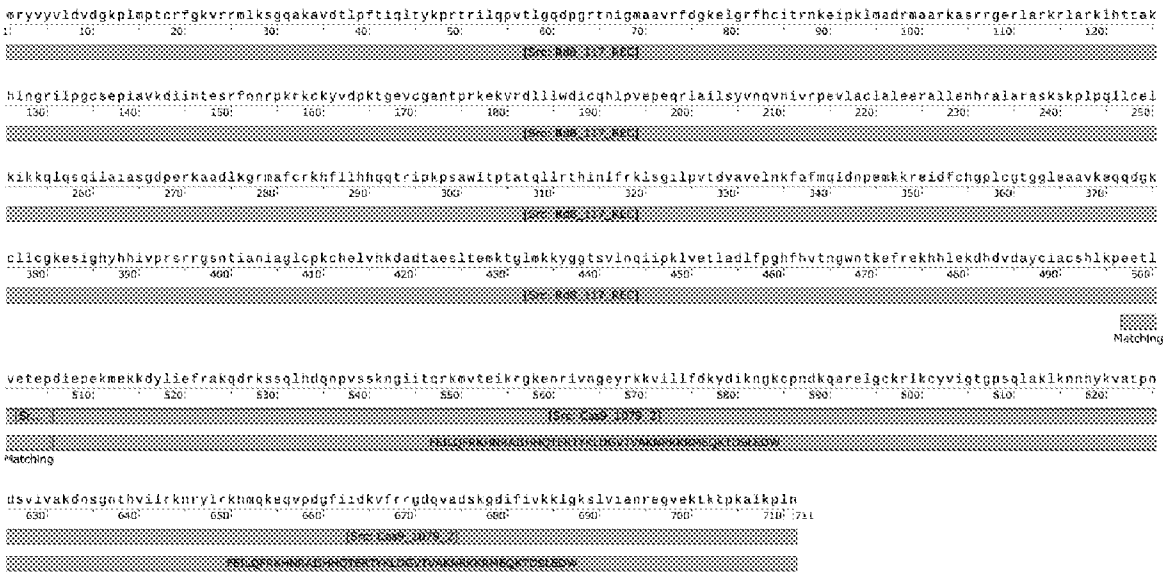


FIG. 54

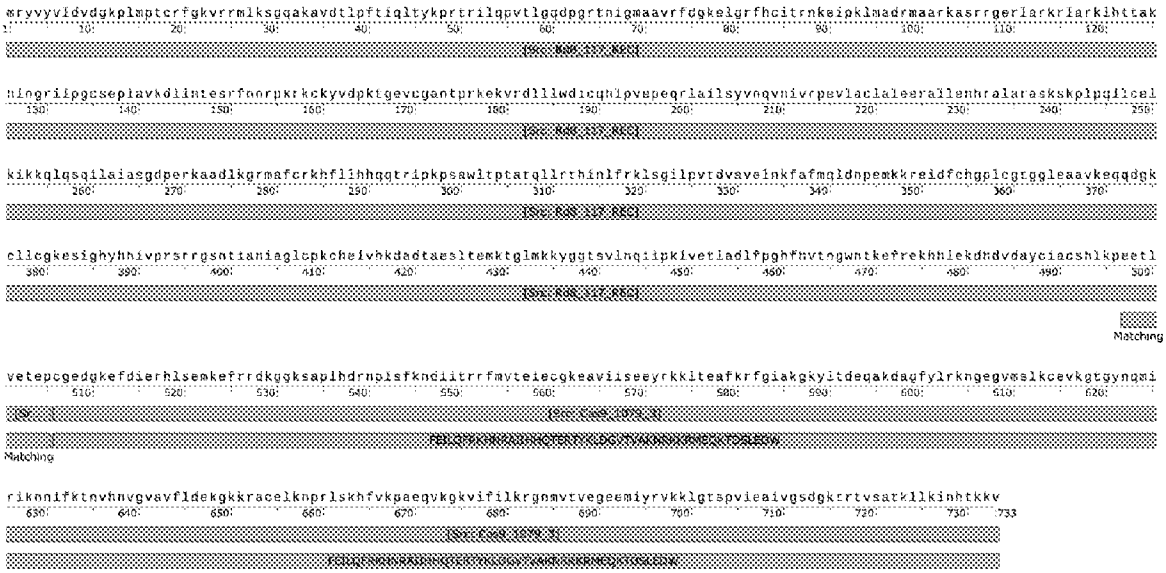


FIG. 55

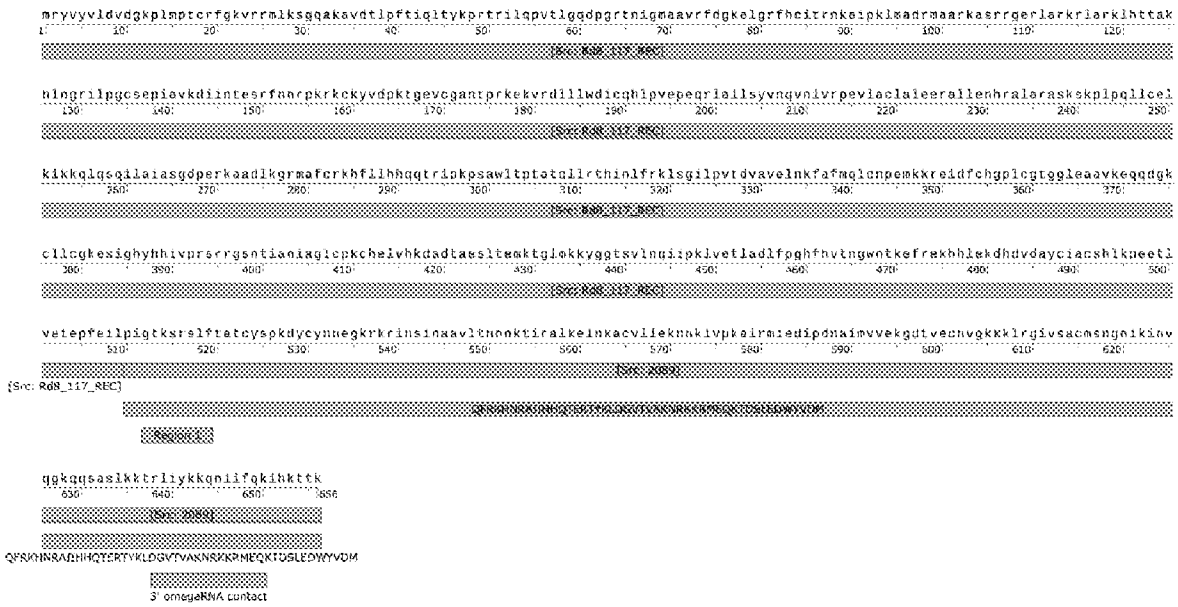


FIG. 56



FIG. 57

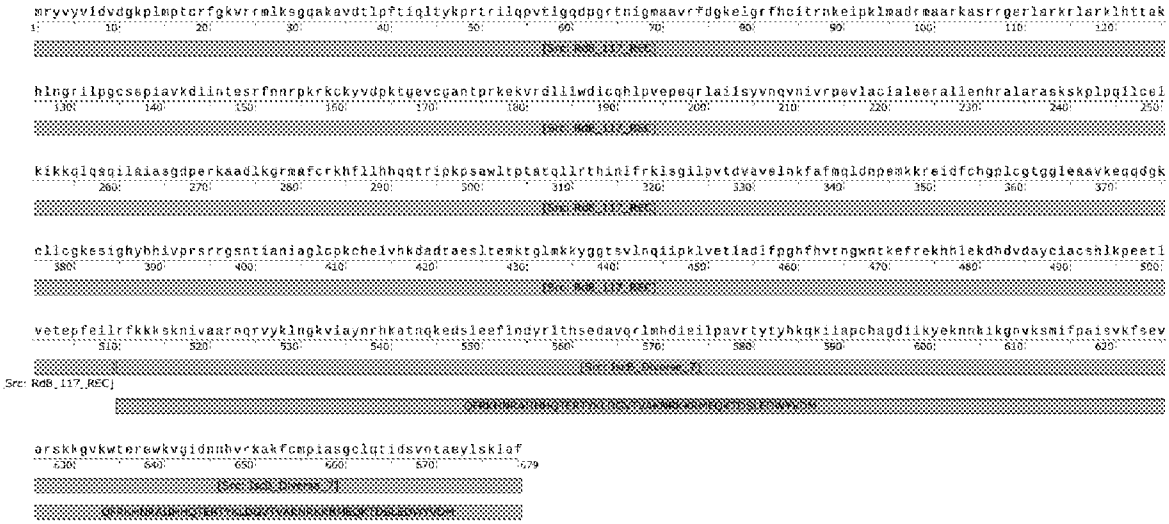


FIG. 58

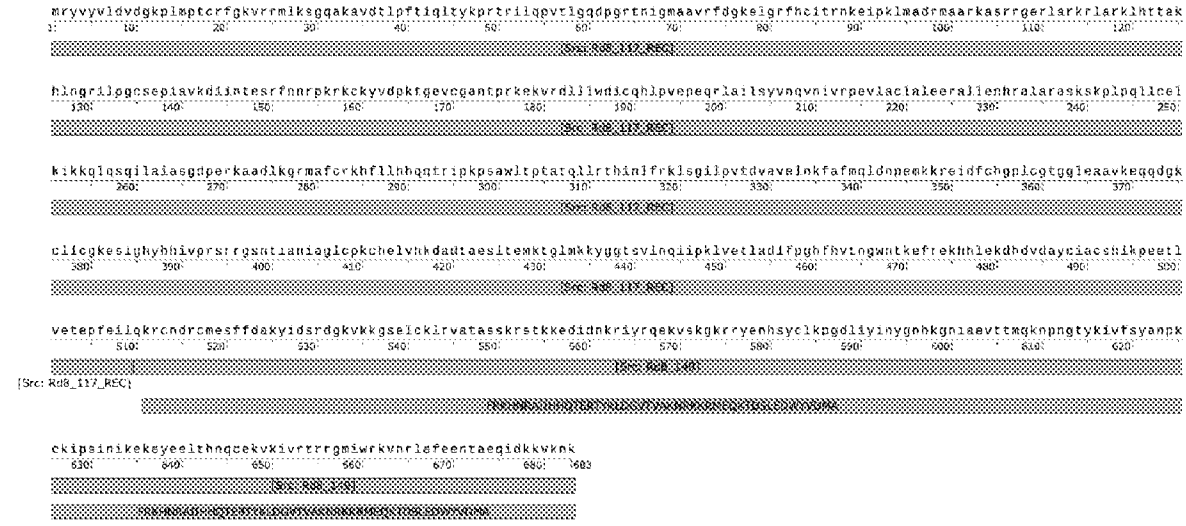


FIG. 59

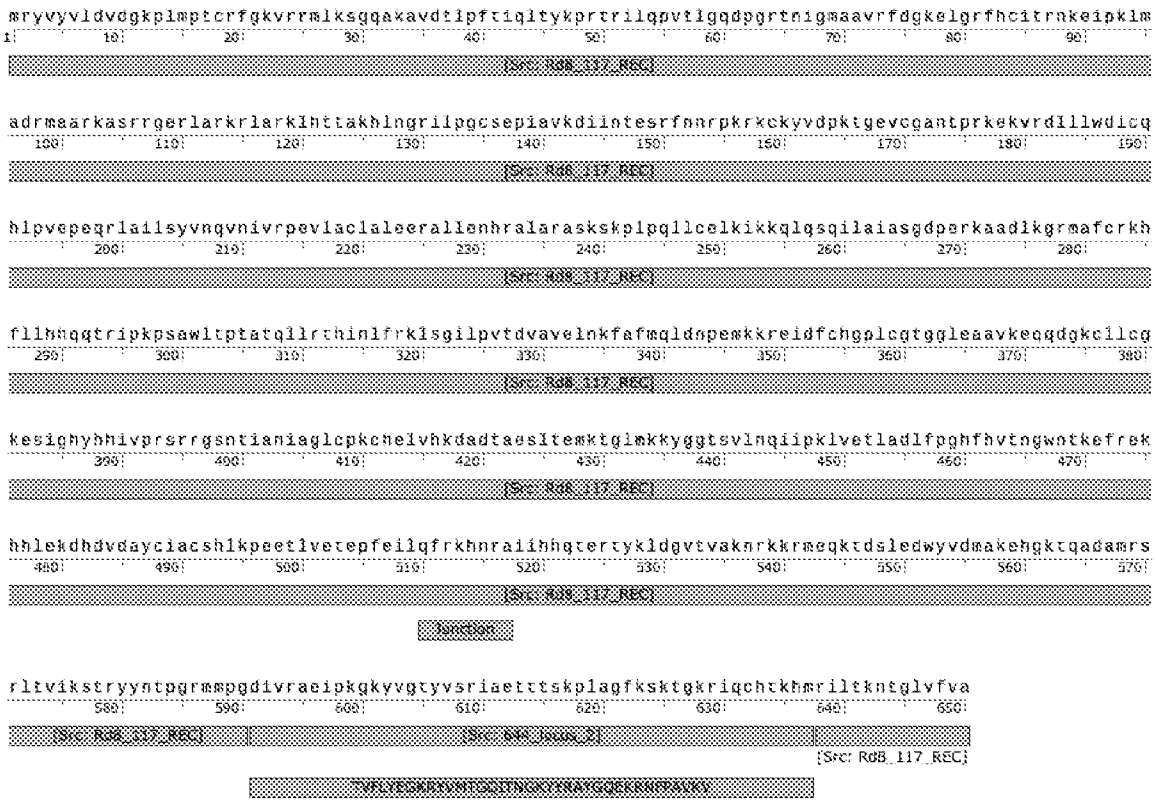

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklm
{Src: Rd8_117_REC}
adrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdillwdicq
{Src: Rd8_117_REC}
hlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
{Src: Rd8_117_REC}
fllhnqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcg
{Src: Rd8_117_REC}
kesighyhhivprsrrgsntianiaglcpkcnelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrek
{Src: Rd8_117_REC}
hhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrs
{Src: Rd8_117_REC}
Junction
rltvikstryyntpgrmmpgdivraeipkgkyvgtyvsriaetttskplagfksktgkriqchtkhmriltkntglvfva
{Src: Rd8_117_REC}
{Src: Rd8_locus_2}
{Src: Rd8_117_REC}
TVPLYEGKPKVMTSQITNGKYRKAYGQEKRNGPAVKV

FIG. 60

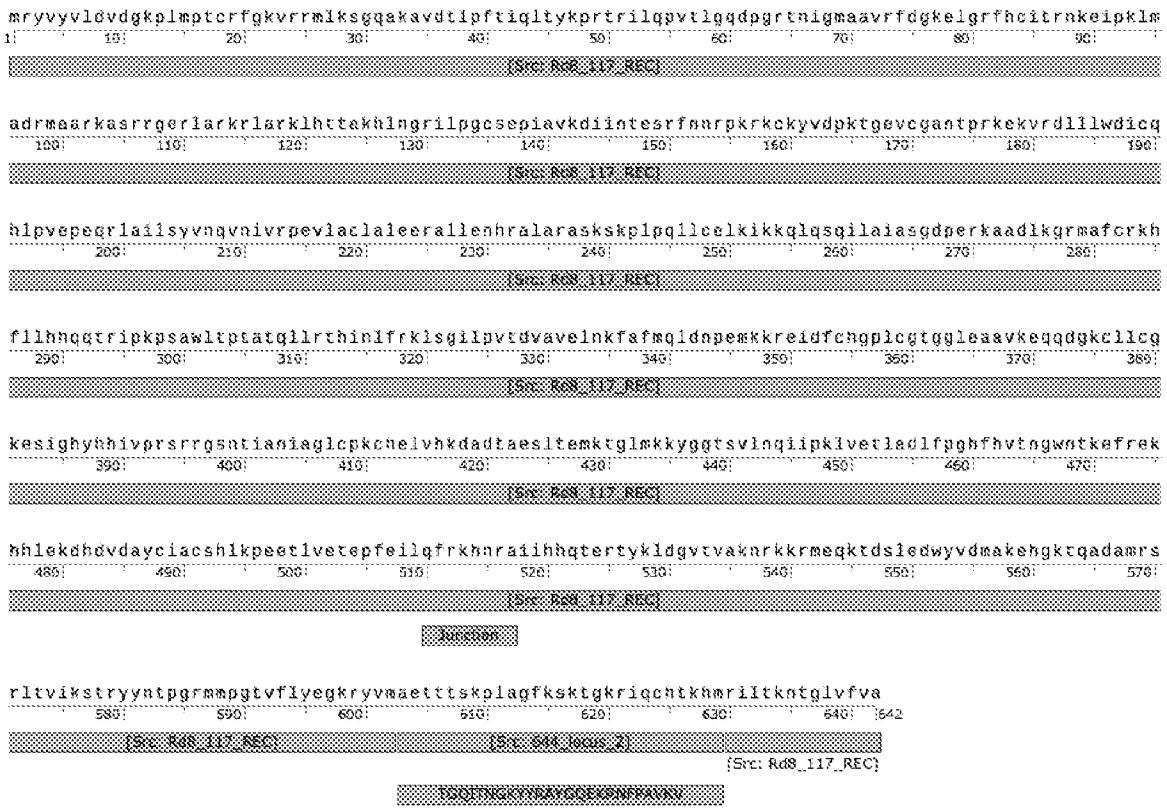

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklm
{Src: Rd8_117_REC}
adrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdillwdicq
{Src: Rd8_117_REC}
hlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
{Src: Rd8_117_REC}
fllhnqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcg
{Src: Rd8_117_REC}
kesighyhhivprsrrgsntianiaglcpkcnelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrek
{Src: Rd8_117_REC}
hhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrs
{Src: Rd8_117_REC}
Junction
rltvikstryyntpgrmmpgtvflyegkryvmaetttskplagfksktgkriqchtkhmriltkntglvfva
{Src: Rd8_117_REC}
{Src: Rd8_locus_2}
{Src: Rd8_117_REC}
TSQITNGKYRKAYGQEKRNGPAVKV

FIG. 61

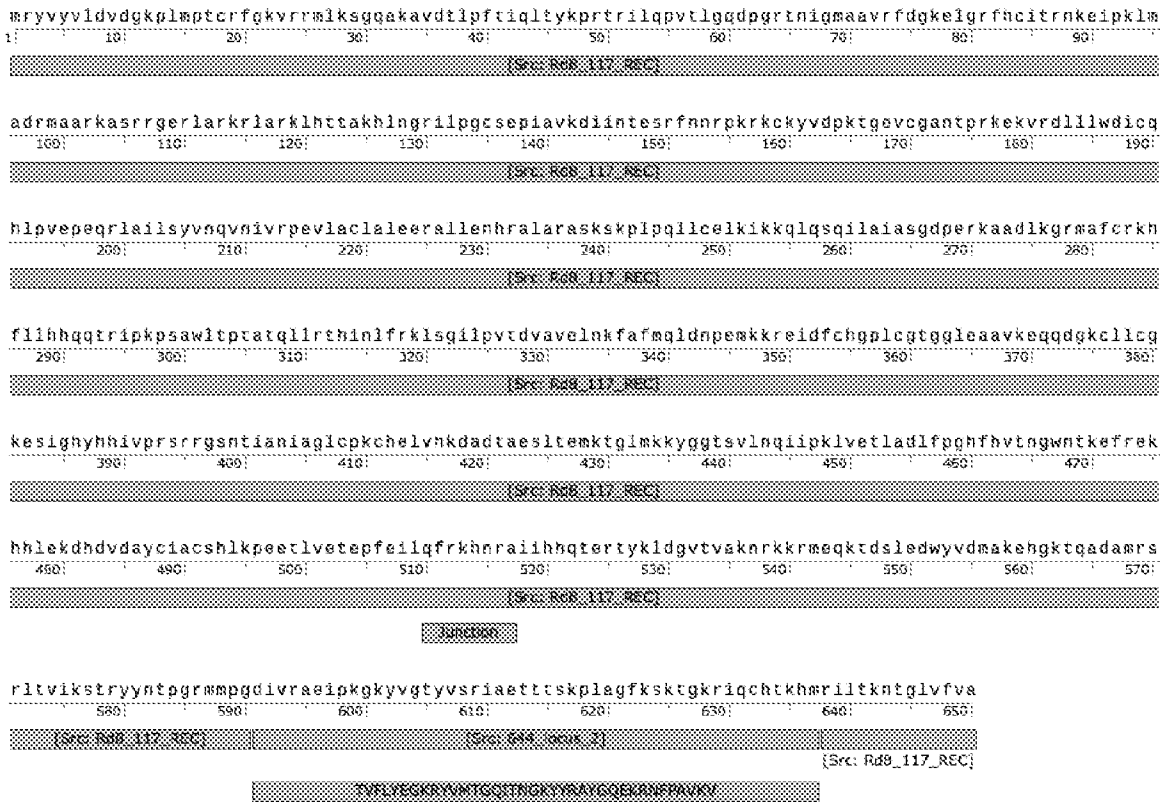

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfncitrnkeipklm
adrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdlllwdicq
hlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqilcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
flihhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcg
kesighyhhivprsrrgsntianiaglcpkchelvnkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrek
hhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnralihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrs
rltvikstryyntpgrmmpgdivraeipkgkyvgtyvsriaetttskplagfksktgkriqchtkhmriltkntglvfva
{Src: Rd8_117_REC}
{Src: 644_orf4_2}
{Src: Rd8_117_REC}

FIG. 62

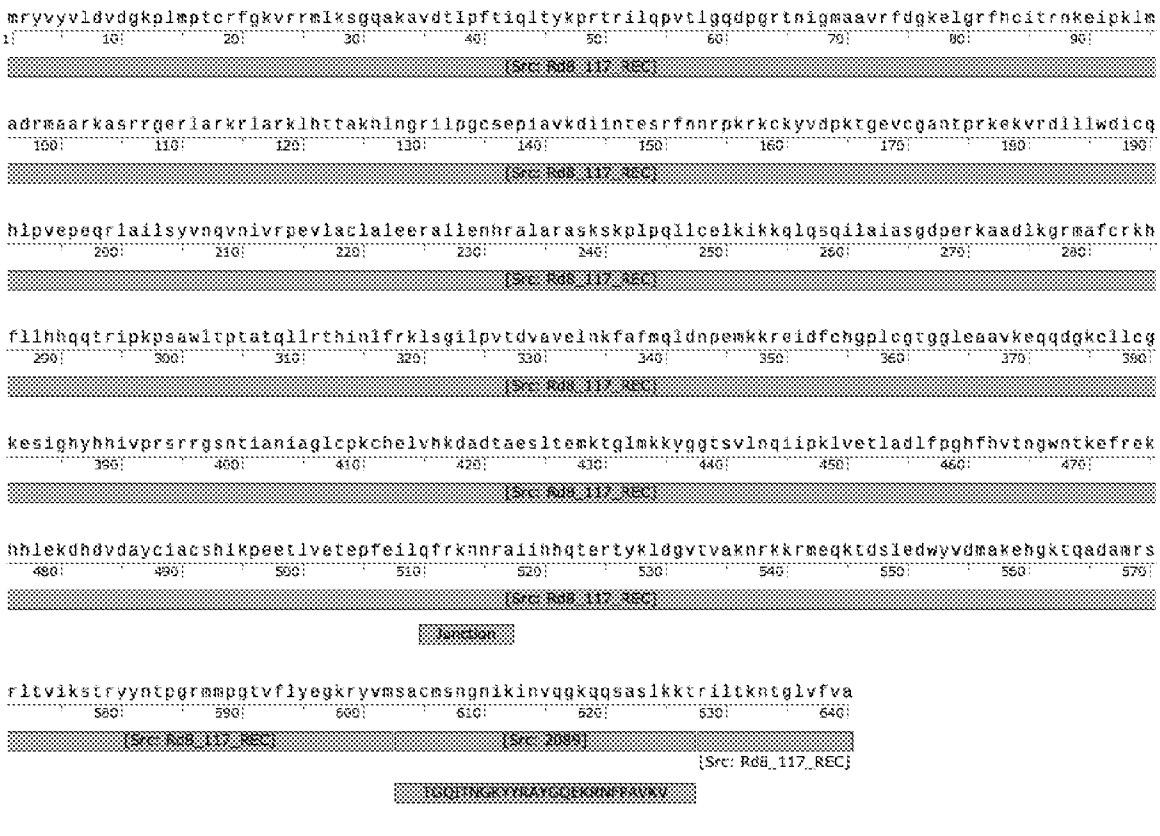

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfncitrnkeipklm
adrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdlllwdicq
hlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqilcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
flihhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcg
kesighyhhivprsrrgsntianiaglcpkchelvnkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrek
hhlekdhdvdayciacshikpeetlvetepfeilqfrkhnralihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrs
rltvikstryyntpgrmmpgtvflyegkryvmsacmsngnikinvqgkqqsaslkktriltkntglvfva
{Src: Rd8_117_REC}
{Src: 2089}
{Src: Rd8_117_REC}

FIG. 63

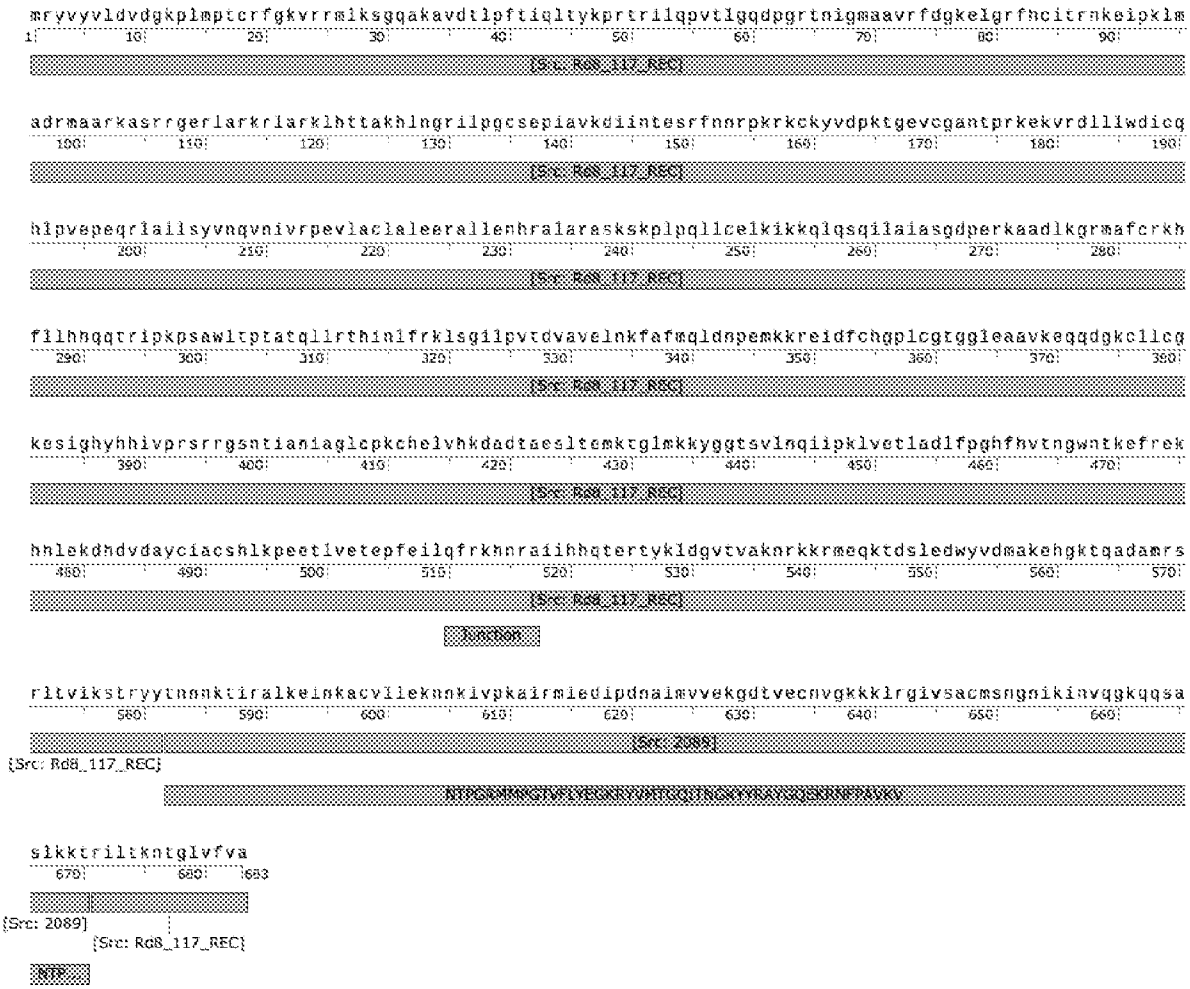

mrvvyvldvdgkplmptcrfgkvrrmiksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklm
1 10 20 30 40 50 60 70 80 90
{Src: Rd8_117_REC}
adrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdlliwdicg
100 110 120 130 140 150 160 170 180 190
{Src: Rd8_117_REC}
hipvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
200 210 220 230 240 250 260 270 280
{Src: Rd8_117_REC}
filhnqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcg
290 300 310 320 330 340 350 360 370 380
{Src: Rd8_117_REC}
kesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwatkefrek
390 400 410 420 430 440 450 460 470
{Src: Rd8_117_REC}
hnlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrs
480 490 500 510 520 530 540 550 560 570
{Src: Rd8_117_REC}
Junction
rltvikstryytnnnktiralkeinkacvileknnkivpkairmiedipdnaimvvekgdtvecnvgkkklrgivsacmsngnikinvggkqqsa
580 590 600 610 620 630 640 650 660
{Src: 2089}
{Src: Rd8_117_REC}
NTPGSMMWGTVLYEGKRYDMTGQTNGKYYRAYQQSKDNPAVKV
slkktriltknqglvfva
670 680 683
{Src: 2089}
{Src: Rd8_117_REC}
NTP

FIG. 64

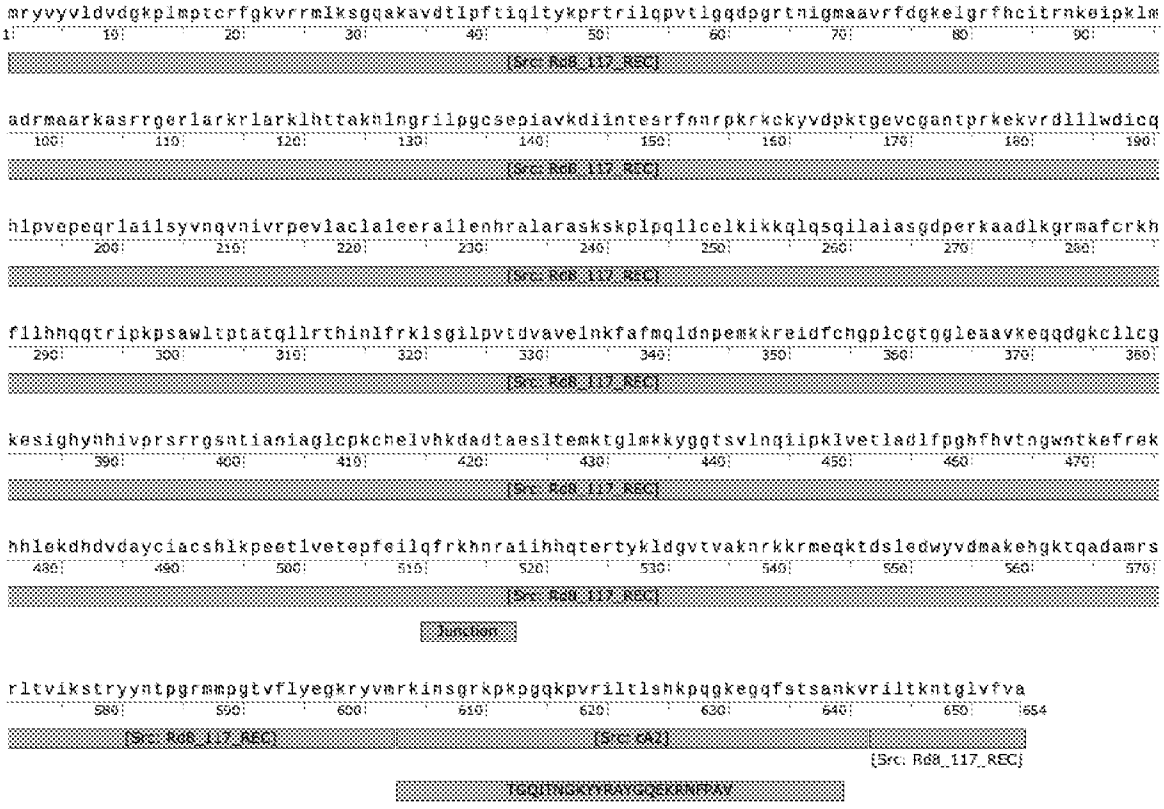

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklm
adrmaarkasrrgerlarkrlarklhttaknlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdlllwdicq
hlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
fllhnqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcg
kesighynhivprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrek
nhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrs
Junction
rltvikstryyntpgrmmpgtvflyegkryvmrkinsgrkpkpgqkpvriltlshkpqgkeqqfstsankvriltkntglvfva
{Src: Rd8_117_REC}
{Src: eA2}
{Src: Rd8_117_REC}
TGQITNSKYYRAYGQEKRNFAV

FIG. 65

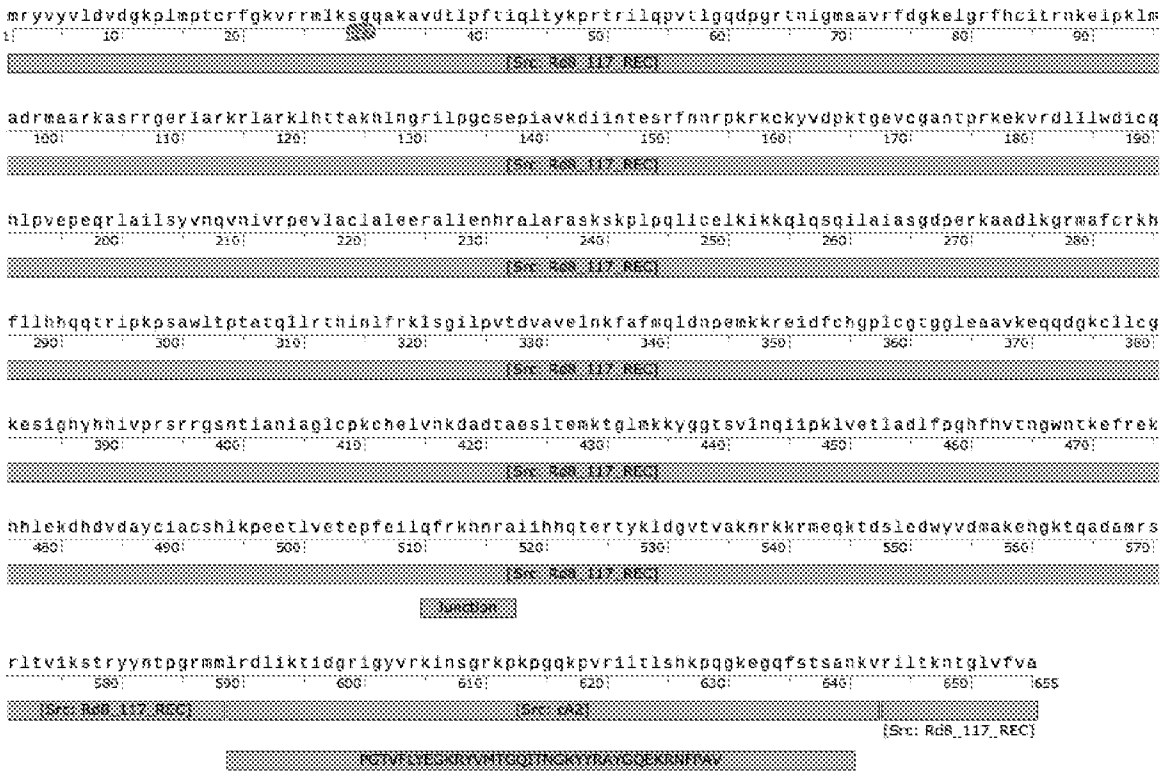

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklm
adrmaarkasrrgerlarkrlarklhttaknlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdlllwdicq
hlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
fllhnqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcg
kesighynhivprsrrgsntianiaglcpkchelvnkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrek
nhlekdhdvdayciacshlkpeetlvetepfeilqfrknnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakengktqadamrs
Junction
rltvikstryyntpgrmmlrdliktidgrigyvrkinsgrkpkpgqkpvriltlshkpqgkeqqfstsankvriltkntglvfva
{Src: Rd8_117_REC}
{Src: eA2}
{Src: Rd8_117_REC}
PGTVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFAV

FIG. 66

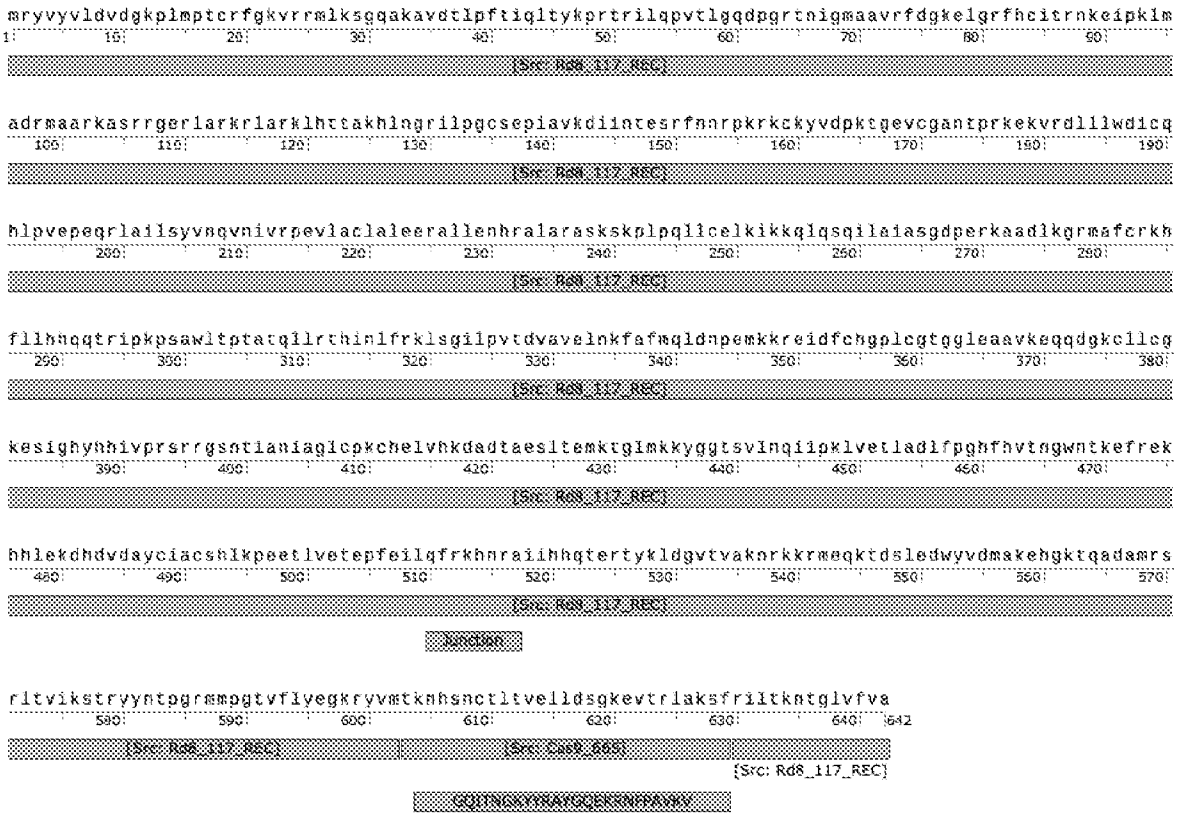

{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
Junction
{Src: Rd8_117_REC}
{Src: Cas9_66}
{Src: Rd8_117_REC}

FIG. 67

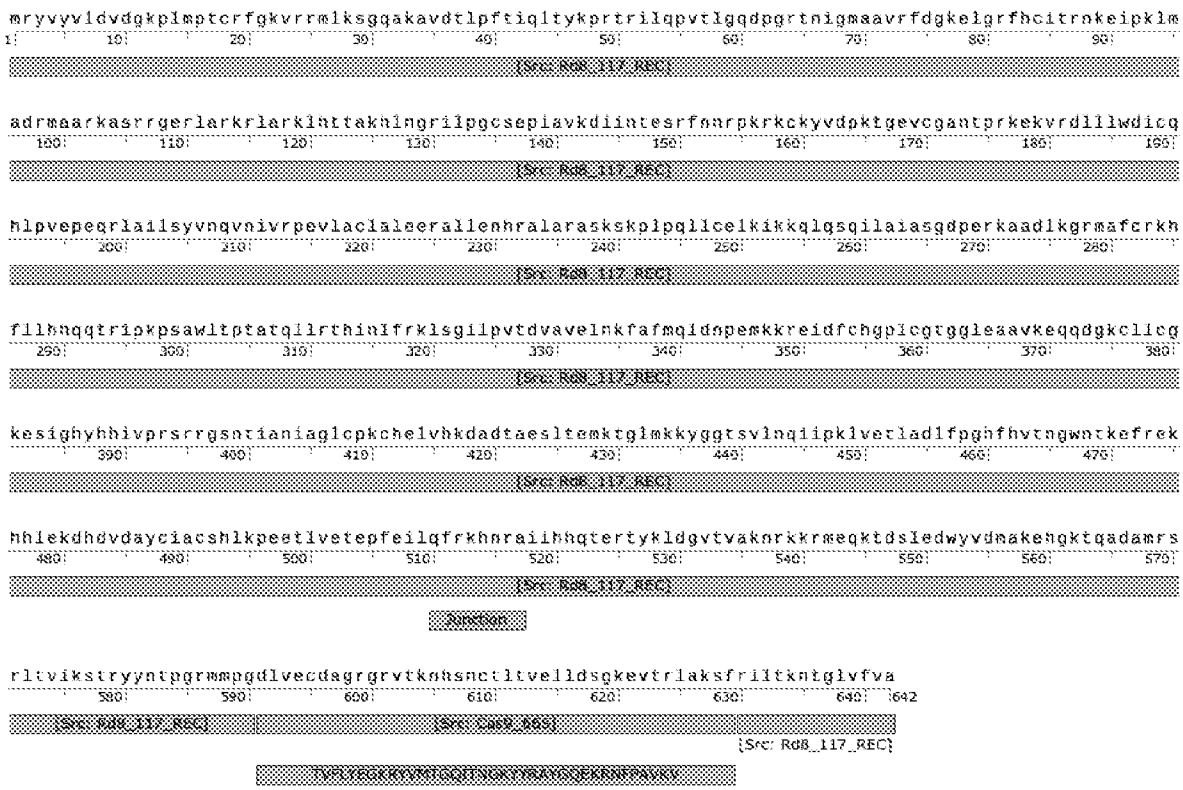

{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
{Src: Rd8_117_REC}
Junction
{Src: Rd8_117_REC}
{Src: Cas9_66}
{Src: Rd8_117_REC}

FIG. 68

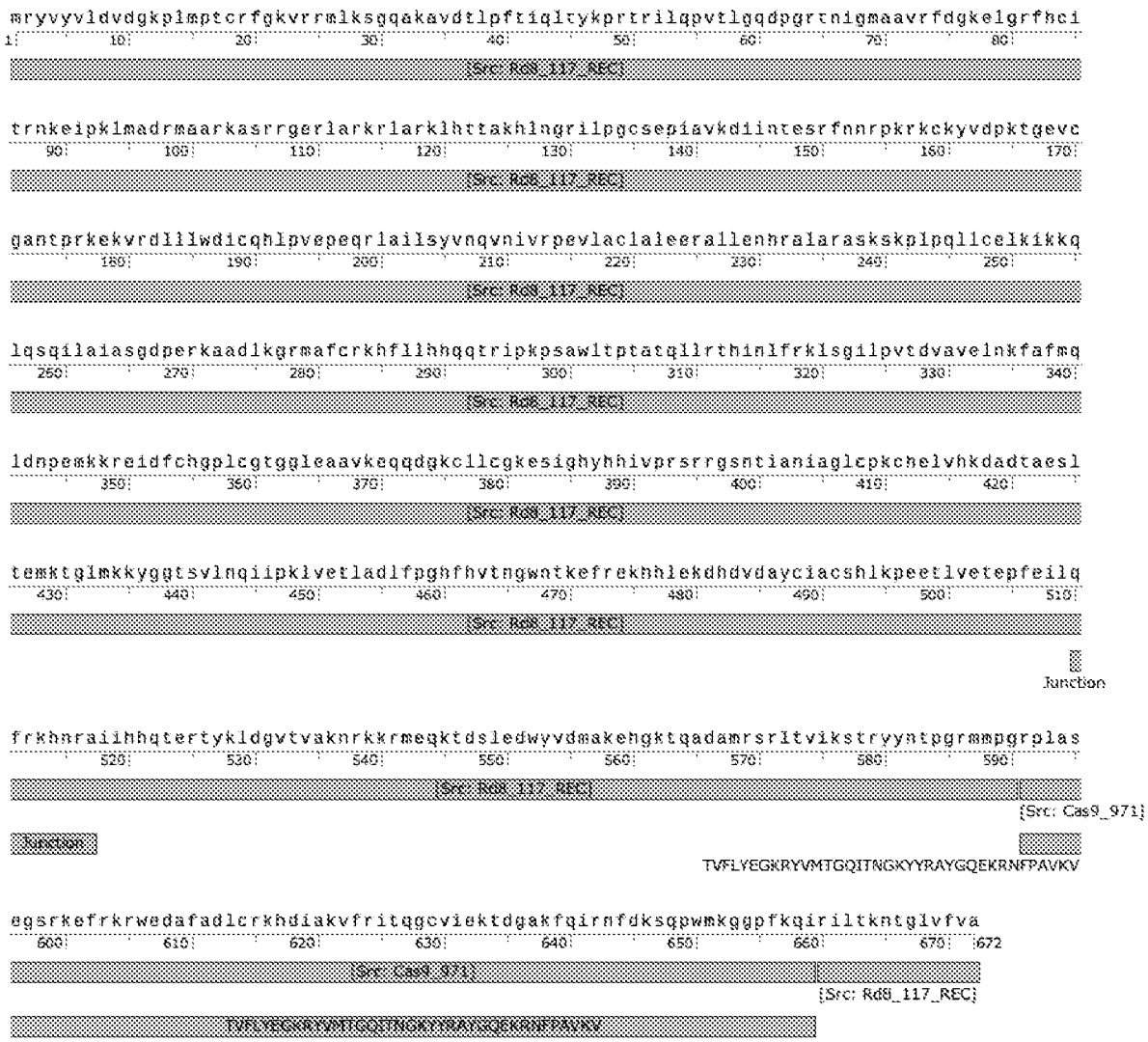

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhci
1 10 20 30 40 50 60 70 80
{Src: Rd8_117_REC}
trnkeipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevc
90 100 110 120 130 140 150 160 170
{Src: Rd8_117_REC}
gantprkekvrdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleerallennralaraskskplpqllcelkikkq
180 190 200 210 220 230 240 250
{Src: Rd8_117_REC}
lqsqilaiasgdperkaadlkgrmafcrkhfllhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmq
260 270 280 290 300 310 320 330 340
{Src: Rd8_117_REC}
ldnpemkkreidfchgplcgtggleaavkeqqdgkcllcgkesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesl
350 360 370 380 390 400 410 420
{Src: Rd8_117_REC}
temktglmkkyggtsvlnqlipklvetladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetlvetepfeilq
430 440 450 460 470 480 490 500 510
{Src: Rd8_117_REC}
Junction
frkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakengktqadamrsritvikstryyntpgrmmpgrplas
520 530 540 550 560 570 580 590
{Src: Rd8_117_REC}
{Src: Cas9_971}
Junction
TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKV
egsrkefrkrwedafadlcrkhdiakvfritqgcviektdgakfqirnfdksqpwmkggpfkqiriltkntglvfva
600 610 620 630 640 650 660 670 672
{Src: Cas9_971}
{Src: Rd8_117_REC}
TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKV

FIG. 69

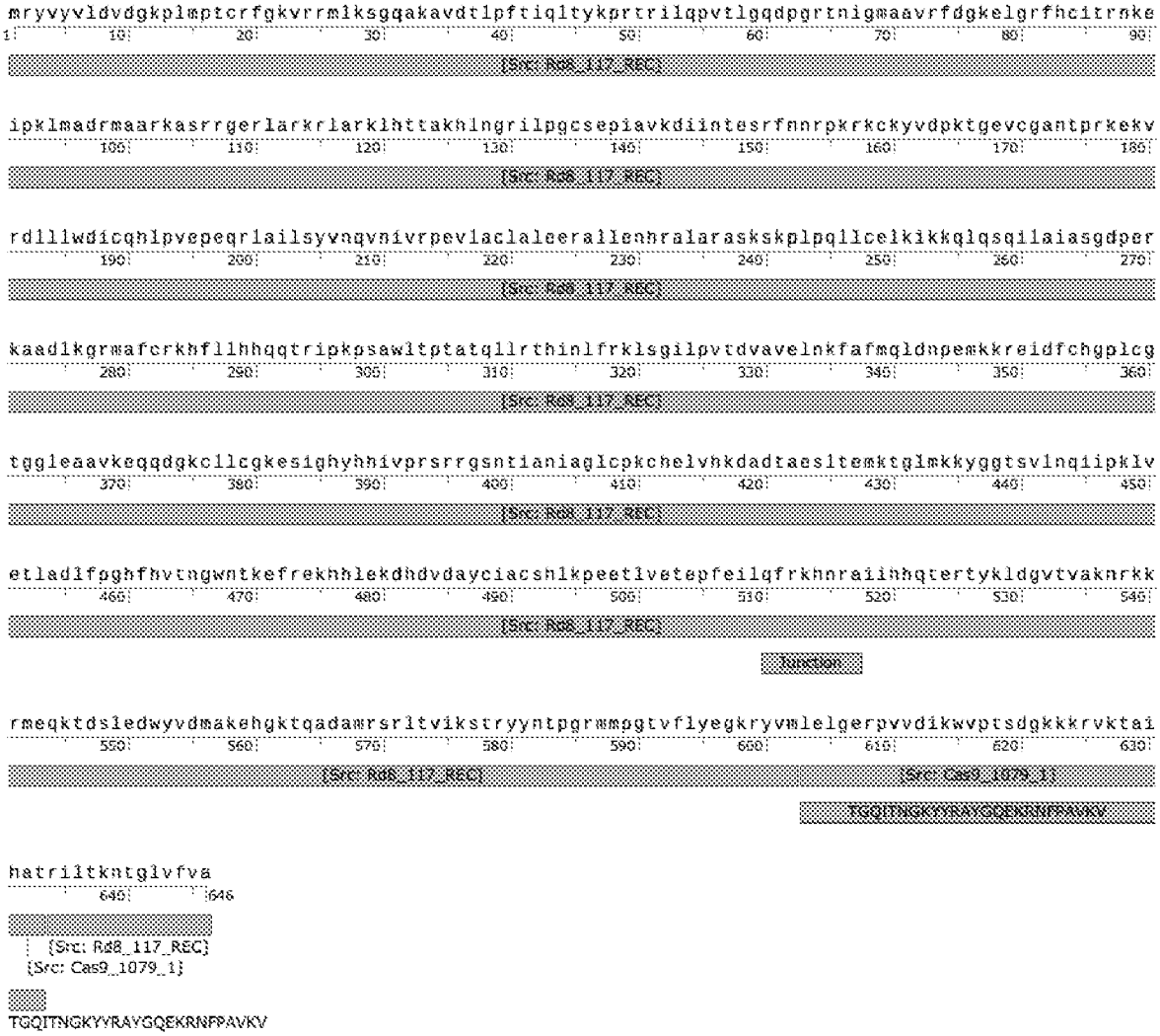

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnke
1 10 20 30 40 50 60 70 80 90
{Src: Rd8_117_REC}
ipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekv
100 110 120 130 140 150 160 170 180
{Src: Rd8_117_REC}
rdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleerallennhralaraskskplpqllcelkikkqlqsqilaiasgdper
190 200 210 220 230 240 250 260 270
{Src: Rd8_117_REC}
kaadlkgrmafcrkhfllhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcg
280 290 300 310 320 330 340 350 360
{Src: Rd8_117_REC}
tggleaavkeqqdgkcllcgkesighyhhnivprsrrgsntianiaglcpkchelvnkdadtaesltemktglmkkyggtsvlnqiipklv
370 380 390 400 410 420 430 440 450
{Src: Rd8_117_REC}
etladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkk
460 470 480 490 500 510 520 530 540
{Src: Rd8_117_REC}
Junction
rmeqktdsledwyvdmakehgktqadamrsrltvikstryyntpgrmmpgtvflyegkryvmlelgerpvvdikwvptsdgkkkrvktai
550 560 570 580 590 600 610 620 630
{Src: Rd8_117_REC}
{Src: Cas9_1079_1}
TGQITNGKYYRAYGQEKRNFPAVKV
hatriltkntglvfva
640 648
{Src: Rd8_117_REC}
{Src: Cas9_1079_1}
TGQITNGKYYRAYGQEKRNFPAVKV

FIG. 70

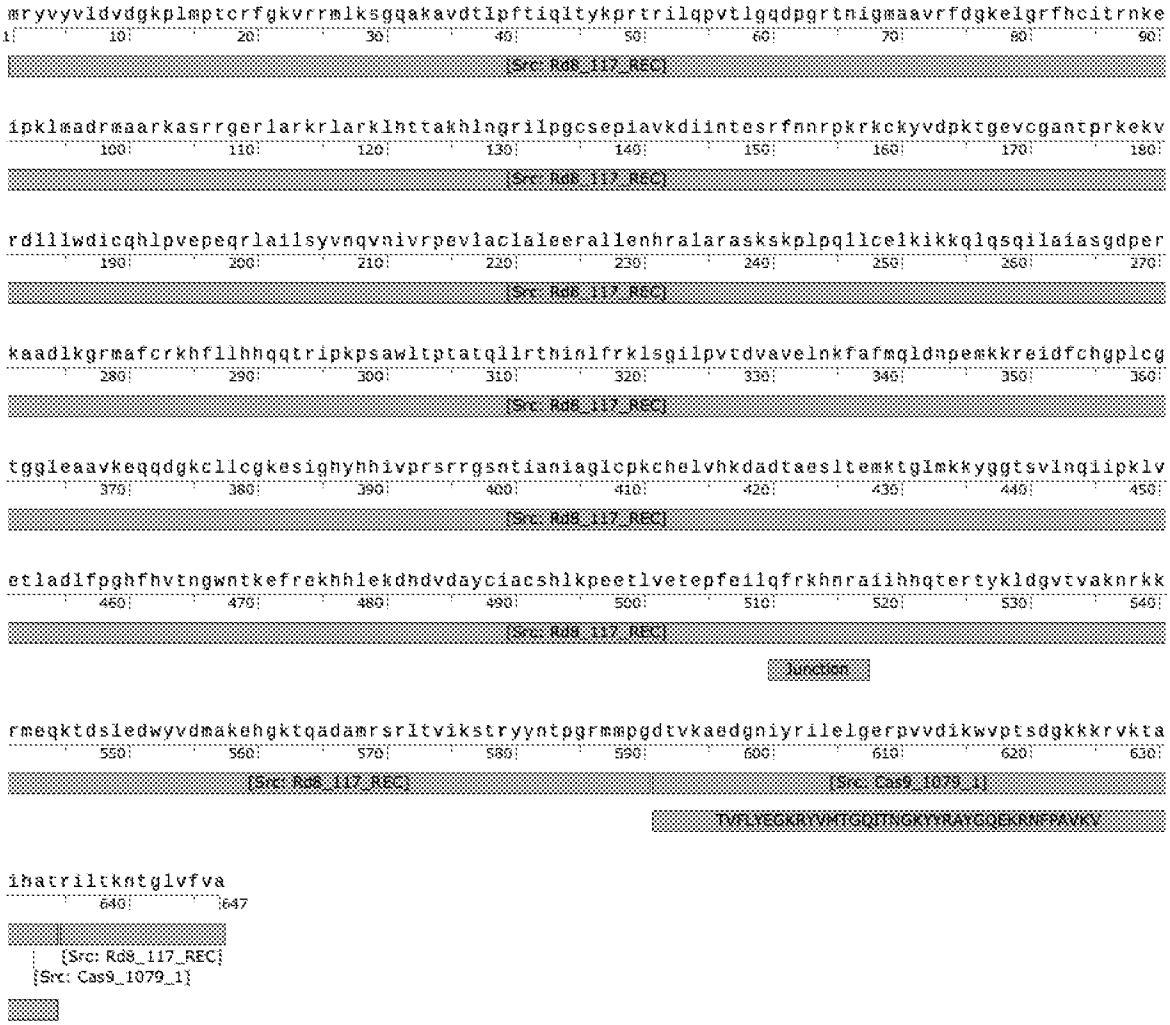

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnke
1 10 20 30 40 50 60 70 80 90
{Src: Rd8_117_REC}
ipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfmnrpkrkckyvdpktgevcgantprkekv
100 110 120 130 140 150 160 170 180
{Src: Rd8_117_REC}
rdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralaraskskplpqllcelkikkqlqsqilaiasgdper
190 200 210 220 230 240 250 260 270
{Src: Rd8_117_REC}
kaadlkgrmafcrkhfllhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcg
280 290 300 310 320 330 340 350 360
{Src: Rd8_117_REC}
tggleaavkeqqdgkcllcgkesighyhhivprsrrgsntianiagicpkchelvhkdadtaesltemktglmkkyggtsvlnqlipklv
370 380 390 400 410 420 430 440 450
{Src: Rd8_117_REC}
etladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetlvetepfeilqfrkhmraiihhqtertykldgvtvaknrkk
460 470 480 490 500 510 520 530 540
{Src: Rd8_117_REC}
Junction
rmeqktdsledwyvdmakehgktqadamrsrltvikstryyntpgrmmpgdtvkaedgniyrilelgerpvvdikwvptsdgkkkrvkta
550 560 570 580 590 600 610 620 630
{Src: Rd8_117_REC}
{Src: Cas9_1079_1}
TVFLYEGKRYWMTGQFTNGKYYRAYCQEKSNFPAVKV
ihatriltkntglvfva
640 647
{Src: Rd8_117_REC}
{Src: Cas9_1079_1}

FIG. 71

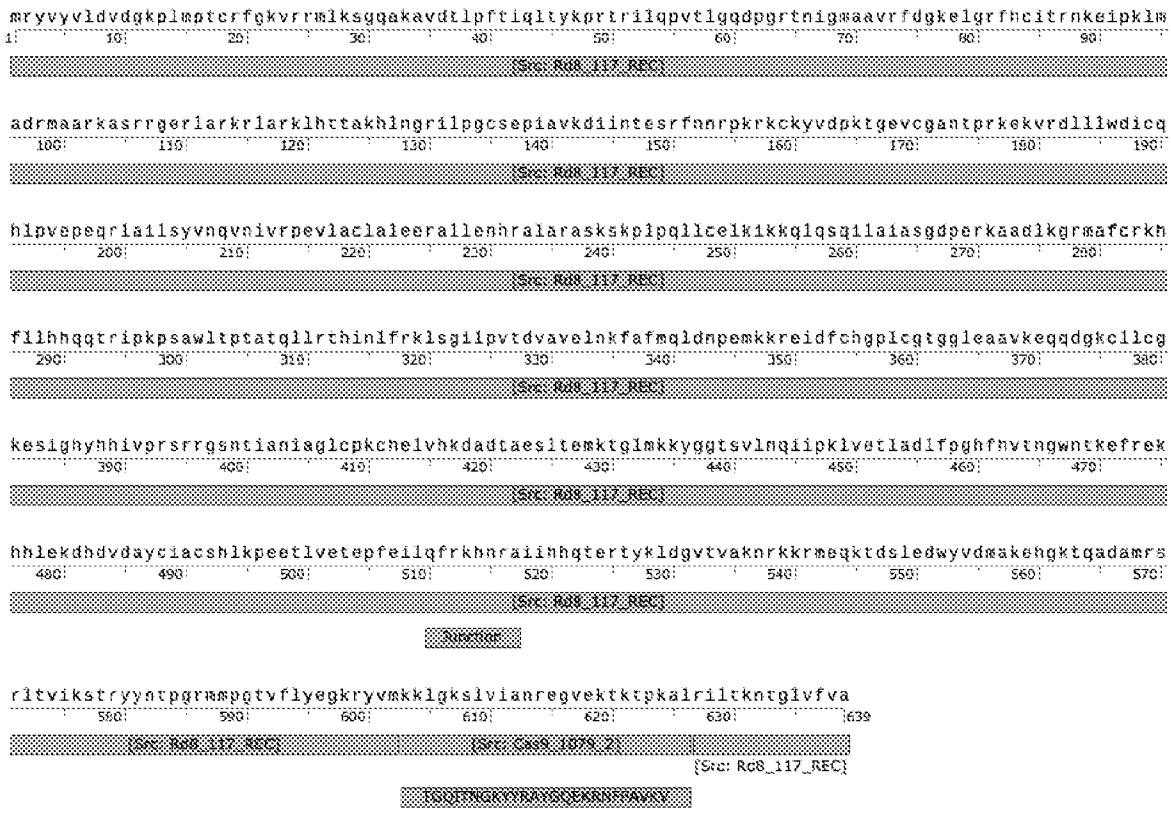

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklm
{Src: Rc8_117_REC}
adrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdllilwdicq
{Src: Rc8_117_REC}
hlpvepeqrlailsyvnqvnivrpevlaclaleerallennhralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkh
{Src: Rc8_117_REC}
flihhqqtripkpsawltptatqllrthinlfrklsgiipvtdvavelnkfafmqldnpemkkreidfchgplcqtggleaavkeqqdgkcllcg
{Src: Rc8_117_REC}
kesighynhivprsrrgsntianlaglcpkcnelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrek
{Src: Rc8_117_REC}
hhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrs
{Src: Rc8_117_REC}
rltvikstryyntpgrmmpgtvflyegkryvmkklgkslvianregvektktpkalriltknrglvfva
{Src: Rc8_117_REC}
{Src: Cas9_1079_2}
{Src: Rc8_117_REC}
TGQITNGKYYRAYSQEKRNFAWYY

FIG. 72

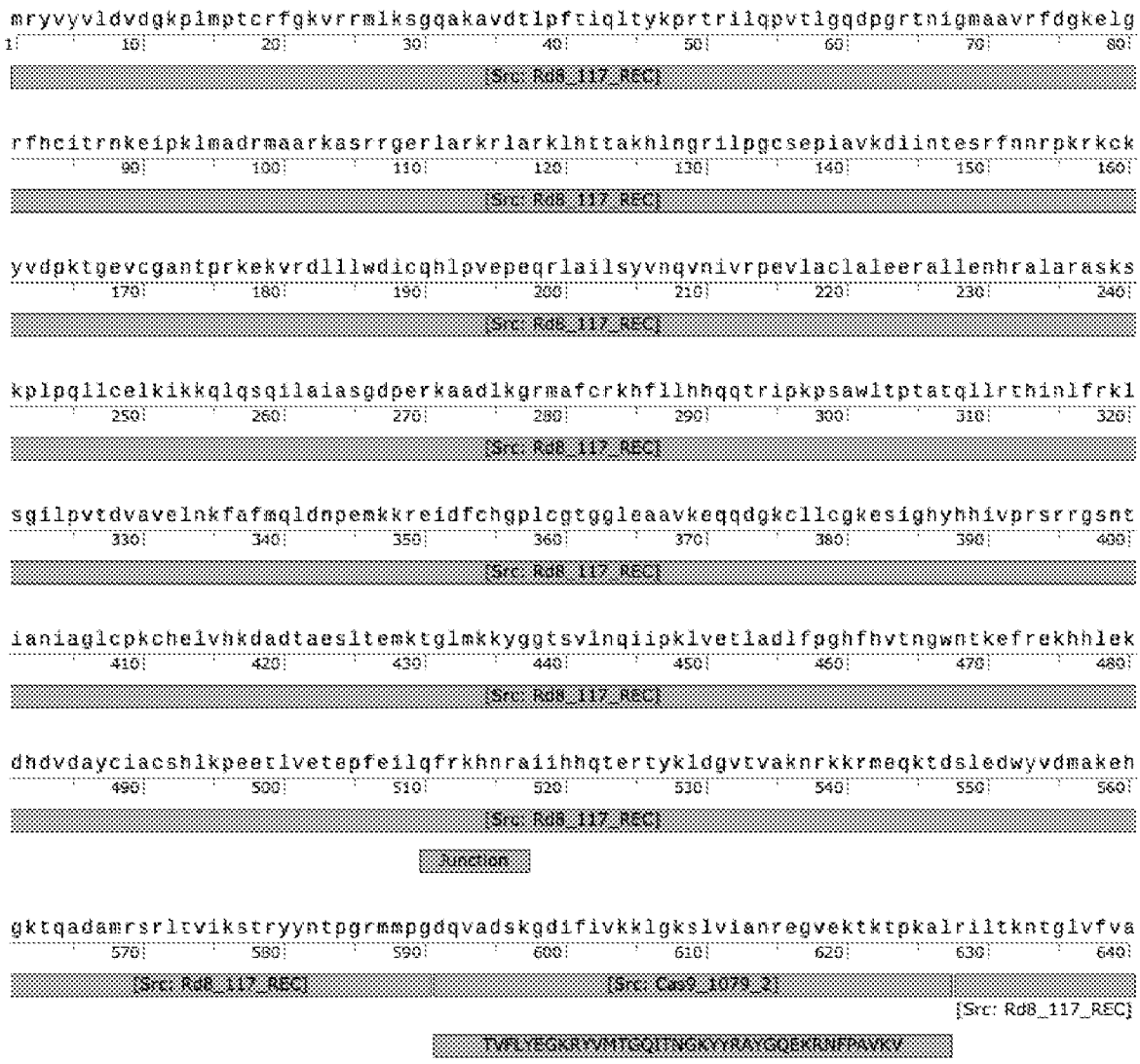

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelg
1 10 20 30 40 50 60 70 80
[Src: Rd8_117_REC]
rfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkck
90 100 110 120 130 140 150 160
[Src: Rd8_117_REC]
yvdpktgevcgantprkekvrdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralarasks
170 180 190 200 210 220 230 240
[Src: Rd8_117_REC]
kplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkhfllhhqqtripkpsawltptatqllrthinlfrkl
250 260 270 280 290 300 310 320
[Src: Rd8_117_REC]
sgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcgkesighyhhivprsrrgsnt
330 340 350 360 370 380 390 400
[Src: Rd8_117_REC]
ianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrekhhlek
410 420 430 440 450 460 470 480
[Src: Rd8_117_REC]
dhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakeh
490 500 510 520 530 540 550 560
[Src: Rd8_117_REC]
Junction
gktqadamrsrltvikstryyntpgrmmpgdqvadskgdifivkklgkslvianregvektktpkalriltkntglvfva
570 580 590 600 610 620 630 640
[Src: Rd8_117_REC]
[Src: Cas9_1079_2]
[Src: Rd8_117_REC]
TVPLYEGKRYVMTCQITNGKYYRAYGQEKRNFPAVKV

FIG. 73

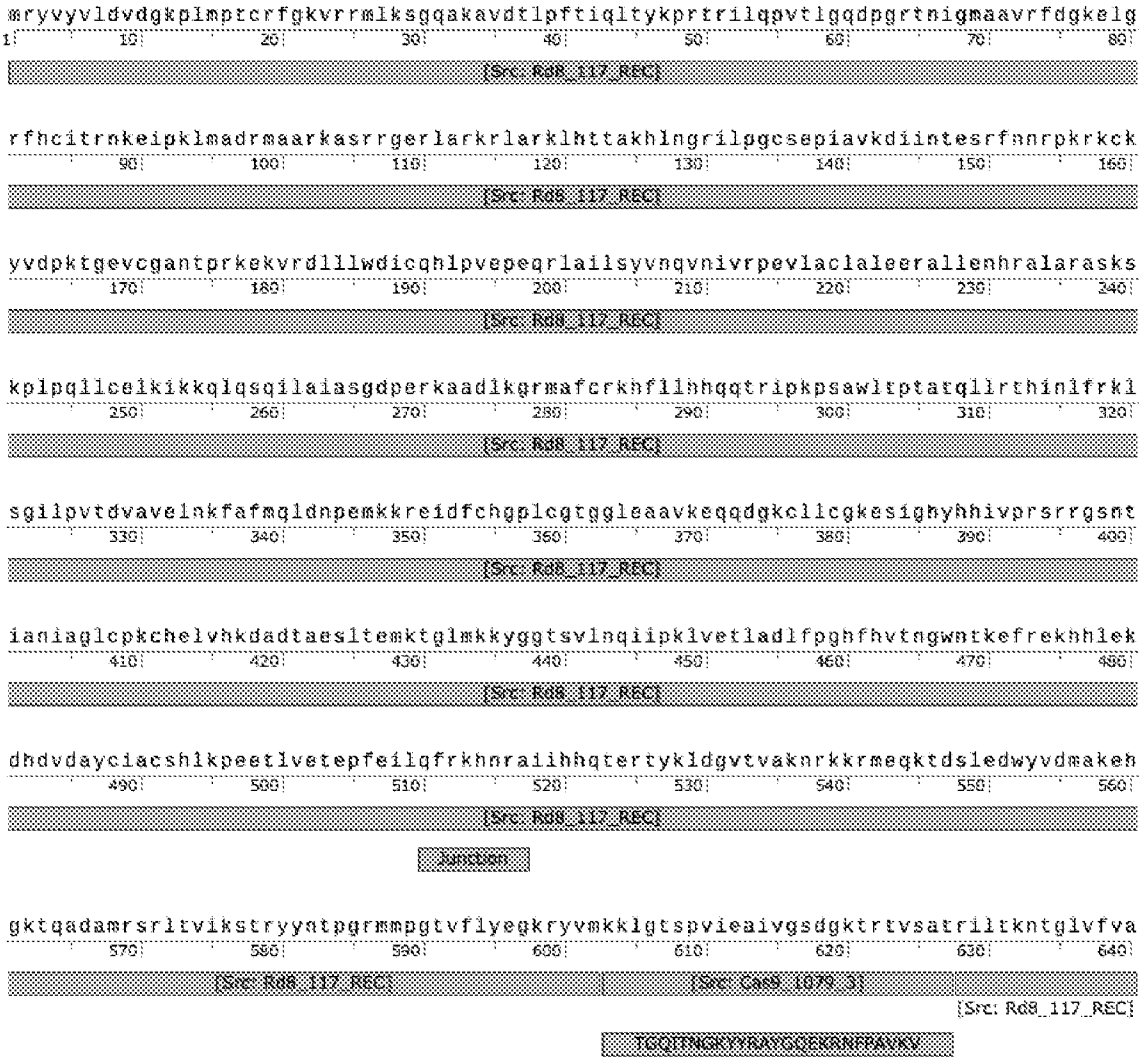

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelg
1 10 20 30 40 50 60 70 80
[Src: Rd8_117_REC]
rfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesrfnnrpkrkck
90 100 110 120 130 140 150 160
[Src: Rd8_117_REC]
yvdpktgevcgantprkekvrdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleerallenhralarasks
170 180 190 200 210 220 230 240
[Src: Rd8_117_REC]
kplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkhfilhhgqtripkpsawltptatqllrthinlfrkl
250 260 270 280 290 300 310 320
[Src: Rd8_117_REC]
sgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgkcllcgkesighyhhivprsrrgsnt
330 340 350 360 370 380 390 400
[Src: Rd8_117_REC]
ianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladlfpghfhvtngwntkefrekhhlek
410 420 430 440 450 460 470 480
[Src: Rd8_117_REC]
dhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakeh
490 500 510 520 530 540 550 560
[Src: Rd8_117_REC]
Junction
gktqadamrsrltvikstryyntpgrmmpgtvflyegkryvmkklgtspvieaivgsdgktrtvsatriltkntglvfva
570 580 590 600 610 620 630 640
[Src: Rd8_117_REC]
[Src: Cas9_1079_3]
[Src: Rd8_117_REC]
TGQITNGKYYRAYSQEKANFPAVKV

FIG. 74

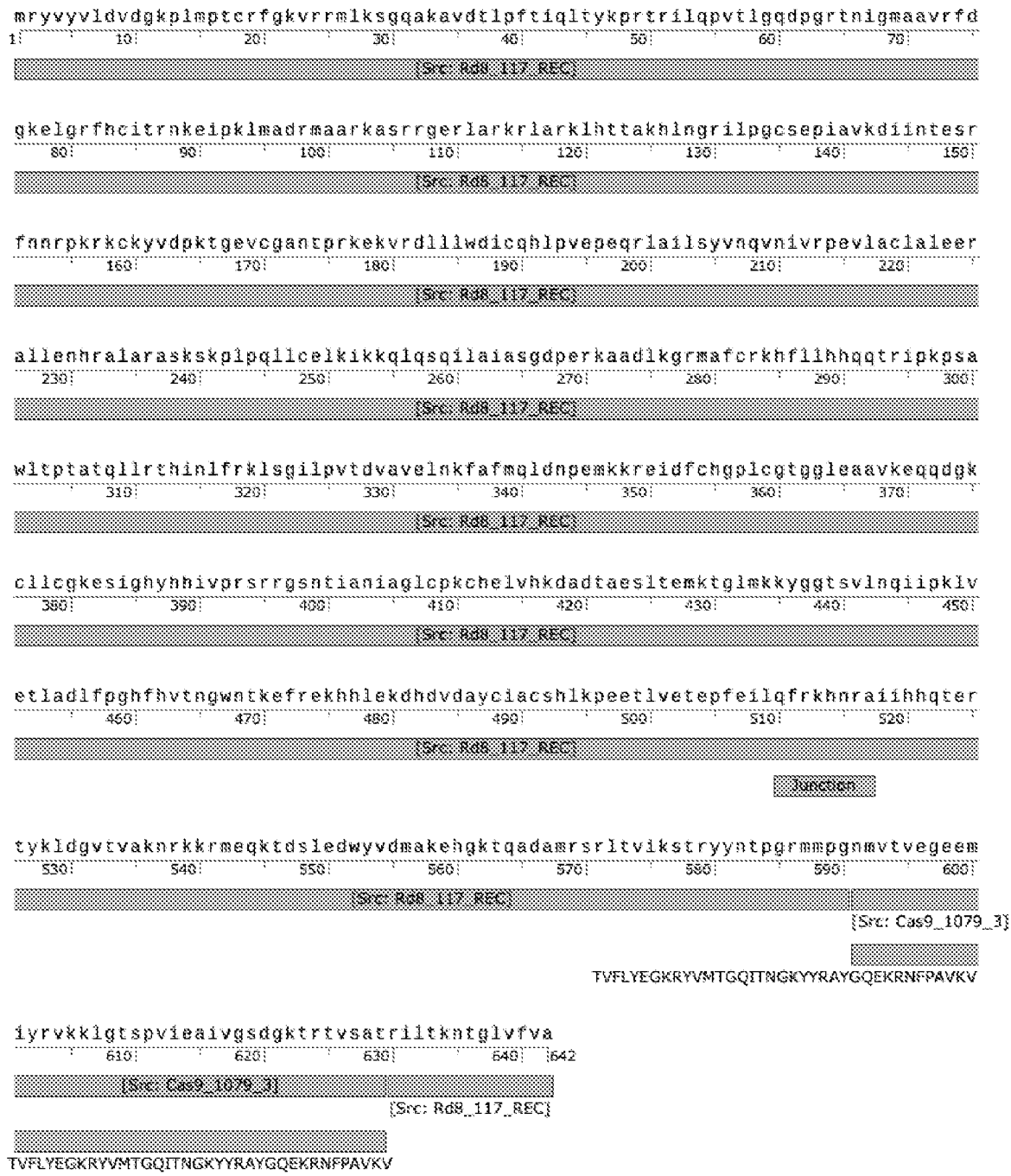

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfd
1 10 20 30 40 50 60 70
{Src: Rd8_117_REC}
gkelgrfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesr
80 90 100 110 120 130 140 150
{Src: Rd8_117_REC}
fnnrpkrkckyvdpktgevcgantprkekvrdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleer
160 170 180 190 200 210 220
{Src: Rd8_117_REC}
allennralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkhflhhqqtripkpsa
230 240 250 260 270 280 290 300
{Src: Rd8_117_REC}
wltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
310 320 330 340 350 360 370
{Src: Rd8_117_REC}
cllcgkesighynhivprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklv
380 390 400 410 420 430 440 450
{Src: Rd8_117_REC}
etladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqter
460 470 480 490 500 510 520
{Src: Rd8_117_REC}
Junction
tykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrsrltvikstryyntpgrmmpgnmvtvegeem
530 540 550 560 570 580 590 600
{Src: Rd8_117_REC}
{Src: Cas9_1079_3}
TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKV
iyrvkklgtspvieaivgsdgktrtvsatriltkntglvfva
610 620 630 640 642
{Src: Cas9_1079_3}
{Src: Rd8_117_REC}
TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKV

FIG. 75

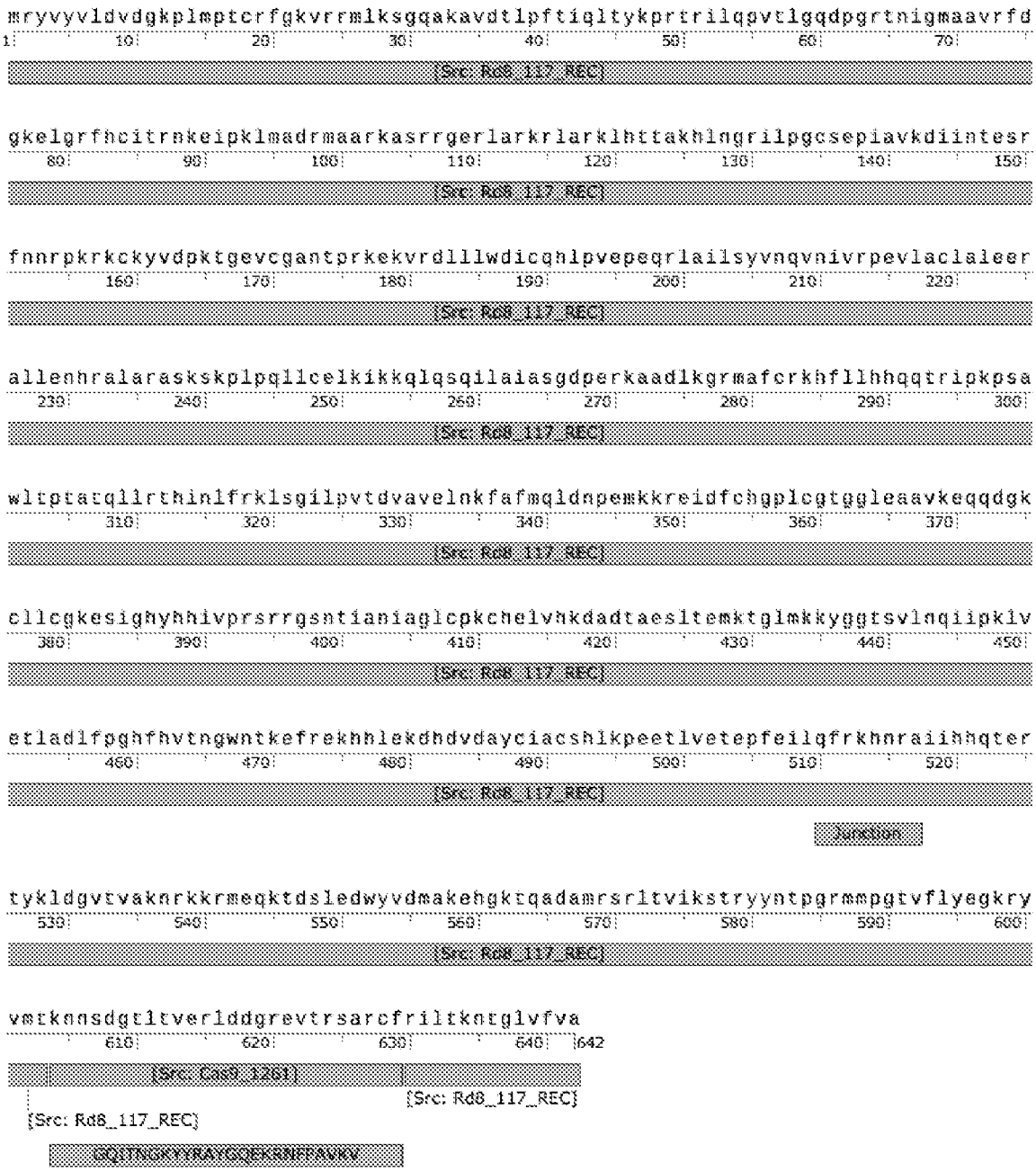

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqgvtlgqdpgrtnigmaavrfd
1 10 20 30 40 50 60 70
[Src: Rd8_117_REC]
gkelgrfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesr
80 90 100 110 120 130 140 150
[Src: Rd8_117_REC]
fnnrpkrkckyvdpktgevcgantprkekvrdlllwdicqhlpvepeqrlailsyvnqvnivrpevlaclaleer
160 170 180 190 200 210 220
[Src: Rd8_117_REC]
allennralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkhfllhhqqtripkpsa
239 240 250 260 270 280 290 300
[Src: Rd8_117_REC]
wltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
310 320 330 340 350 360 370
[Src: Rd8_117_REC]
cllcgkesighyhhivprsrrgsntianiaglcpkchelvnkdadtaesltemktglmkkyggtsvlnqiipklv
380 390 400 410 420 430 440 450
[Src: Rd8_117_REC]
etladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqter
460 470 480 490 500 510 520
[Src: Rd8_117_REC]
Junction
tykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrsrltvikstryyntpgrmmpgtvflyegkry
530 540 550 560 570 580 590 600
[Src: Rd8_117_REC]
vmtknnsdgtltverlddgrevtrsarcfriltkntglvfva
610 620 630 640 642
[Src: Cas9_1261]
[Src: Rd8_117_REC]
[Src: Rd8_117_REC]
GQITNGKYYRAYGQEKRNFFAVKV

FIG. 76

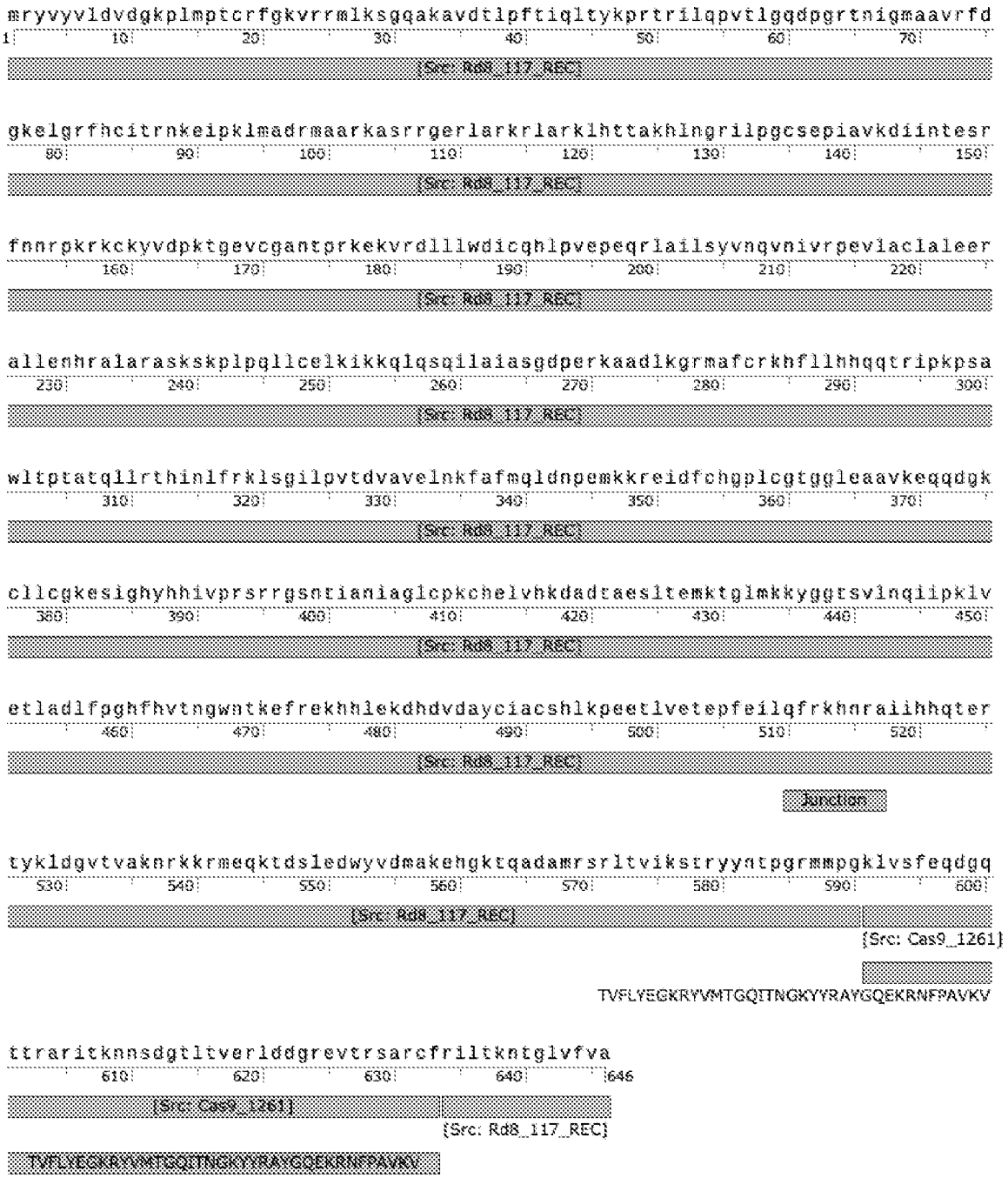

mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfd
1 10 20 30 40 50 60 70
{Src: Rd8_117_REC}
gkelgrfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttakhlngrilpgcsepiavkdiintesr
80 90 100 110 120 130 140 150
{Src: Rd8_117_REC}
fnnrpkrkckyvdpktgevcgantprkekvrdlllwdicghlpvepeqrlailsyvnqvnivrpevlaclaleer
160 170 180 190 200 210 220
{Src: Rd8_117_REC}
allennralaraskskplpqllcelkikkqlqsqilaiasgdperkaadlkgrmafcrkhfllhhqqtripkpsa
230 240 250 260 270 280 290 300
{Src: Rd8_117_REC}
wltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
310 320 330 340 350 360 370
{Src: Rd8_117_REC}
cllcgkesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvinqiipklv
380 390 400 410 420 430 440 450
{Src: Rd8_117_REC}
etladlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetlvetepfeilqfrkhnraiihhqter
460 470 480 490 500 510 520
{Src: Rd8_117_REC}
Junction
tykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrsrltvikstryynpgrmmpgklvsfeqdgq
530 540 550 560 570 580 590 600
{Src: Rd8_117_REC}
{Src: Cas9_1261}
TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKV
ttraritknnsdgtltverlddgrevtrsarcfriltkntglvfva
610 620 630 640 646
{Src: Cas9_1261}
{Src: Rd8_117_REC}
TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKV

FIG. 77


mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttak
[Src: Rd8_117_REC]
hlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdliiwdicqhlpvepeqrlaiisyvnqvnivrpevlaclaleerailenhralaraskskplpqllcei
[Src: Rd8_117_REC]
kikkqlqsqilaiasgdperkaadlkgrmafcrkhflihhqqtripkpsawltptatqllrthinifrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
[Src: Rd8_117_REC]
cllcgkesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladifpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetl
[Src: Rd8_117_REC]
vetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrsrltvikstryyntpgrmmpgkivsfeqdgttraritknnsdgtltveriddgre
[Src: Rd8_117_REC]
[Src: Cas9_1261]
Junction
vtrsarcfriltkntglvfva
[Src: Cas9_1261]
[Src: Rd8_117_REC]
TVFLYEGKRYVMTGQITNGKYYRAYGQEKRNFPAVKV

FIG. 78


mryvyvldvdgkplmptcrfgkvrrmlksgqakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklmadrmaarkasrrgerlarkrlarklhttak
[Src: Rd8_117_REC]
hlngrilpgcsepiavkdiintesrfnnrpkrkckyvdpktgevcgantprkekvrdliiwdicqhlpvepeqrlaiisyvnqvnivrpevlaclaleerailenhralaraskskplpqllcei
[Src: Rd8_117_REC]
kikkqlqsqilaiasgdperkaadlkgrmafcrkhflihhqqtripkpsawltptatqllrthinifrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
[Src: Rd8_117_REC]
cllcgkesighyhhivprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetladifpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetl
[Src: Rd8_117_REC]
vetepfeilqfrkhnraiihhqtertykldgvtvaknrkkrmeqktdsledwyvdmakehgktqadamrsrltvikstryyntpgrmmpgtvflyegkryvmtvratgnfavtpknqkqvsimyk
[Src: Rd8_117_REC]
[Src: Chlor!sc8]
Junction
ycriltkntglvfva
[Src: Rd8_117_REC]
[Src: Chlor!sc8]
GQITNGKYYRAYGQEKRNFPAVKV

FIG. 79

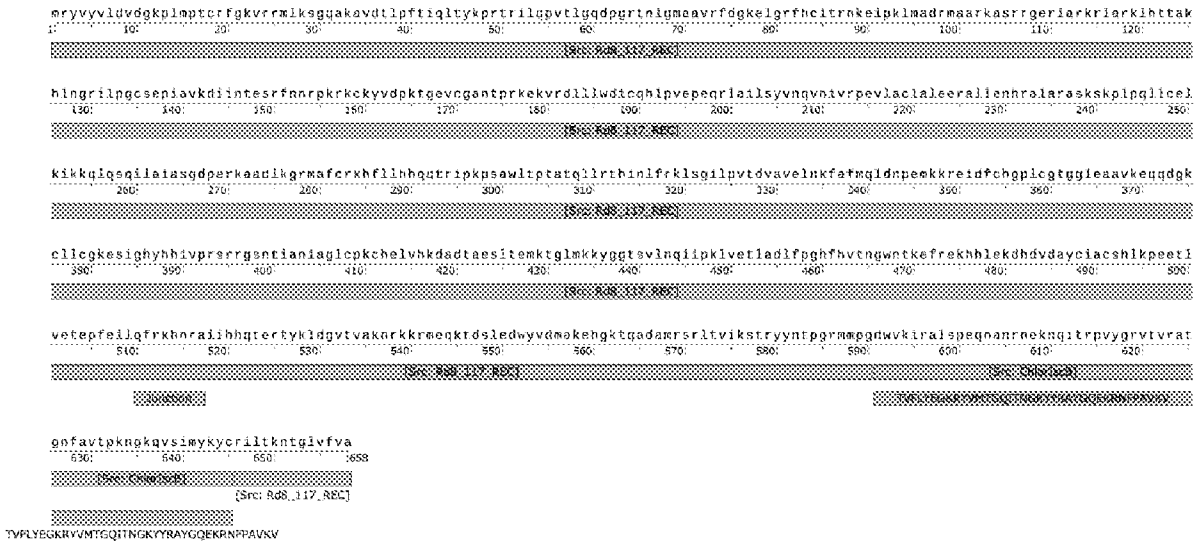

[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]

FIG. 80


[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]
[Src: RdB_117_REC]

FIG. 81

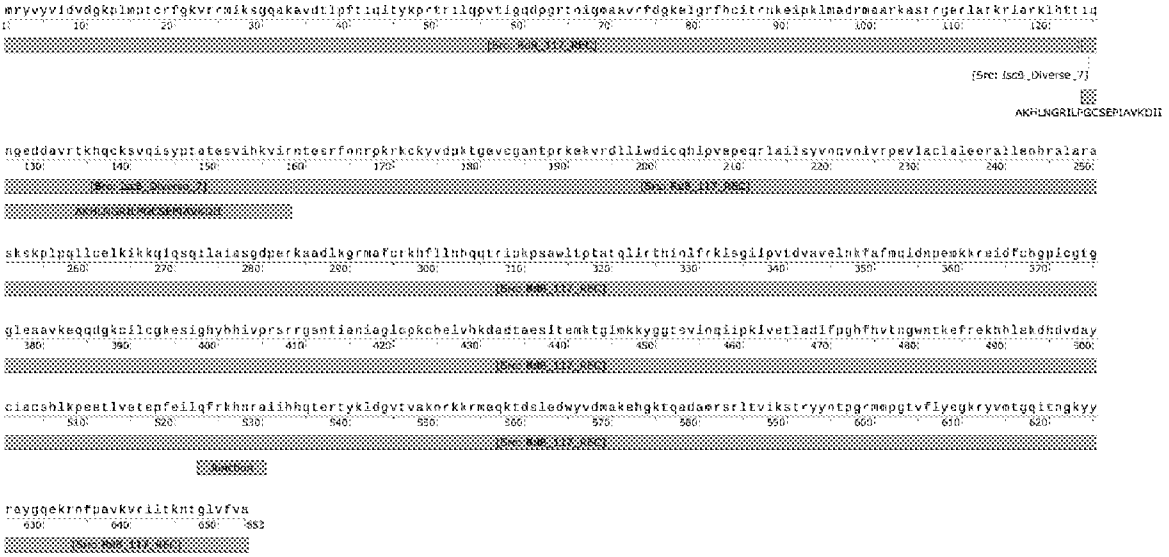

[Src: R08_117_REC]
[Src: Isc8_Diverse_7]
AKFLNGRILPGCSEPIAVKDII
[Src: R08_117_REC]
[Src: R08_117_REC]
[Src: R08_117_REC]
[Src: R08_117_REC]

FIG. 82

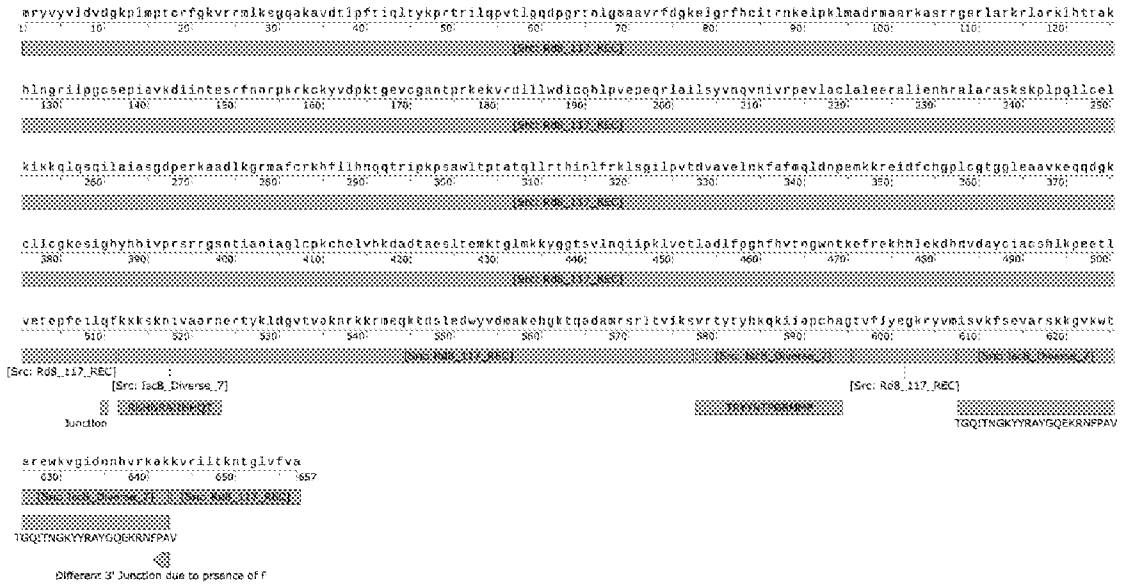

[Src: R08_117_REC]
[Src: Isc8_Diverse_7]
Junction
[Src: R08_117_REC]
TGQITNGKYYRAYSQEKRNFPAV
TGQITNGKYYRAYSQEKRNFPAV
Different 3' Junction due to presence of f

FIG. 83



FIG. 84



FIG. 85



FIG. 86

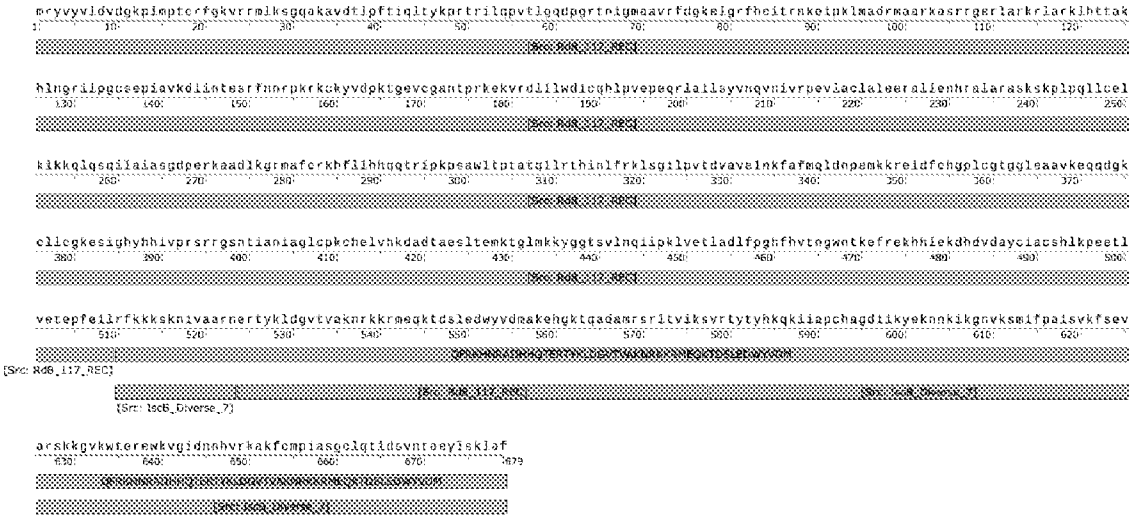


FIG. 87



FIG. 88

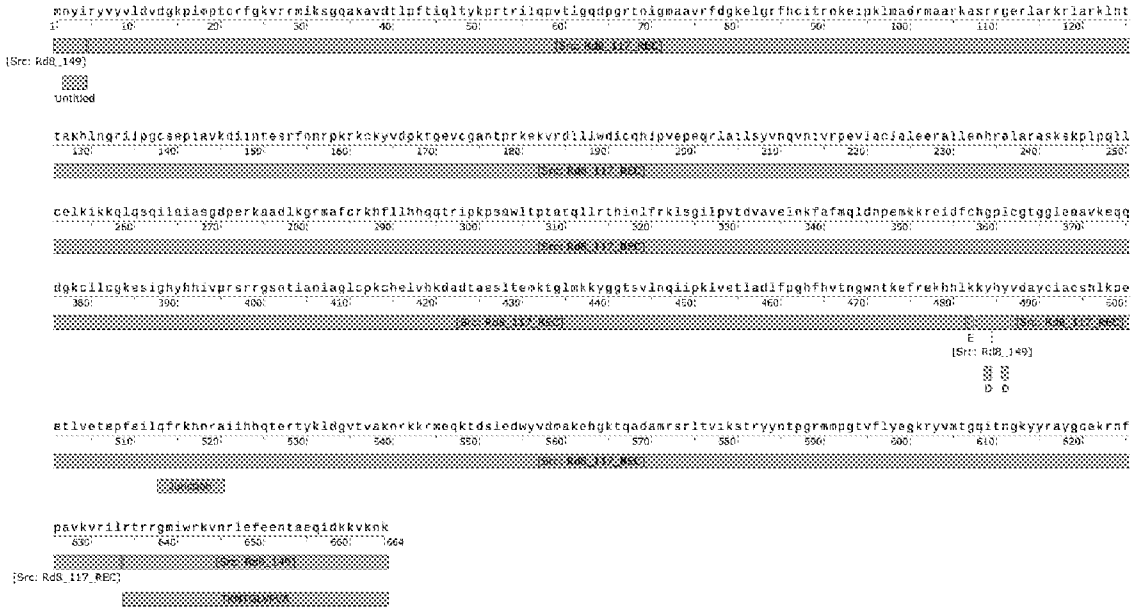


FIG. 89



FIG. 90

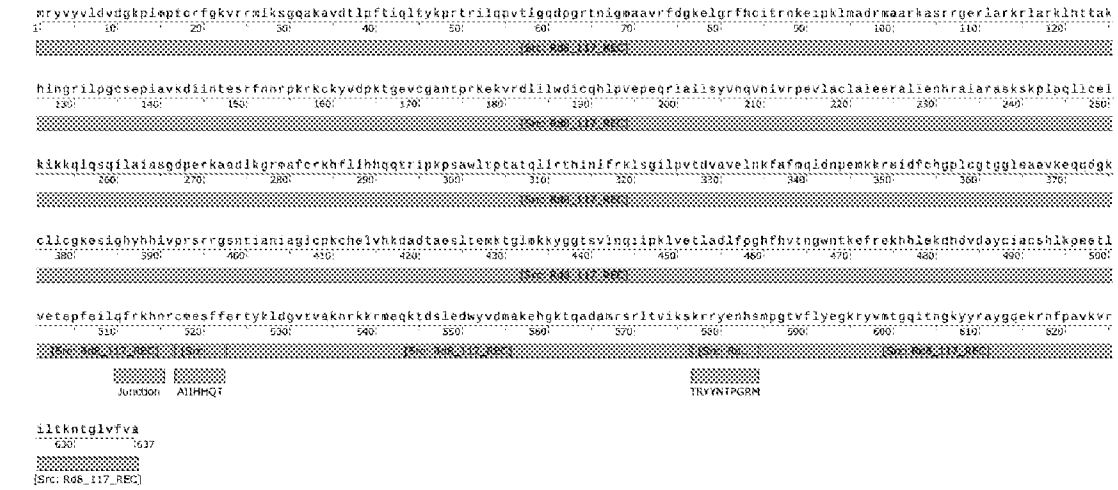


FIG. 91

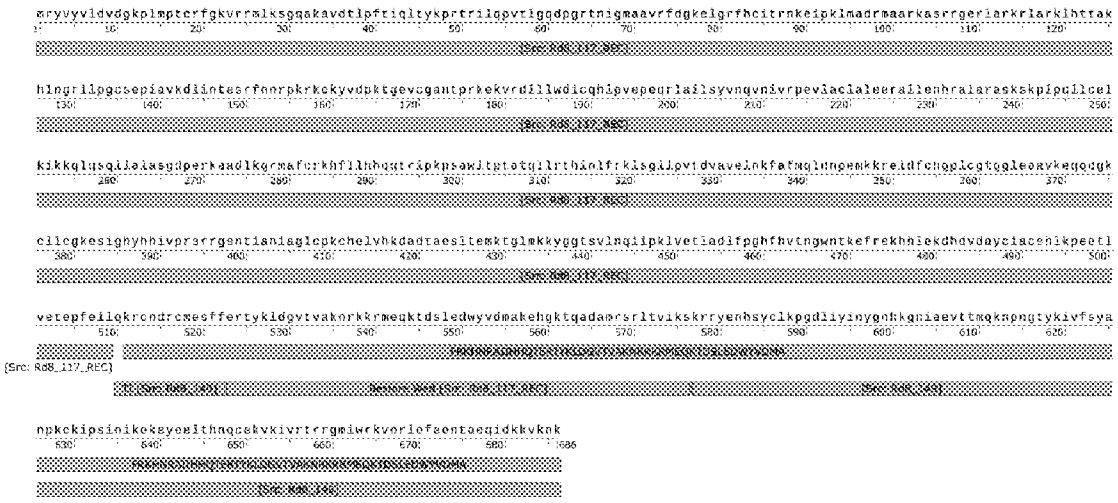


FIG. 92

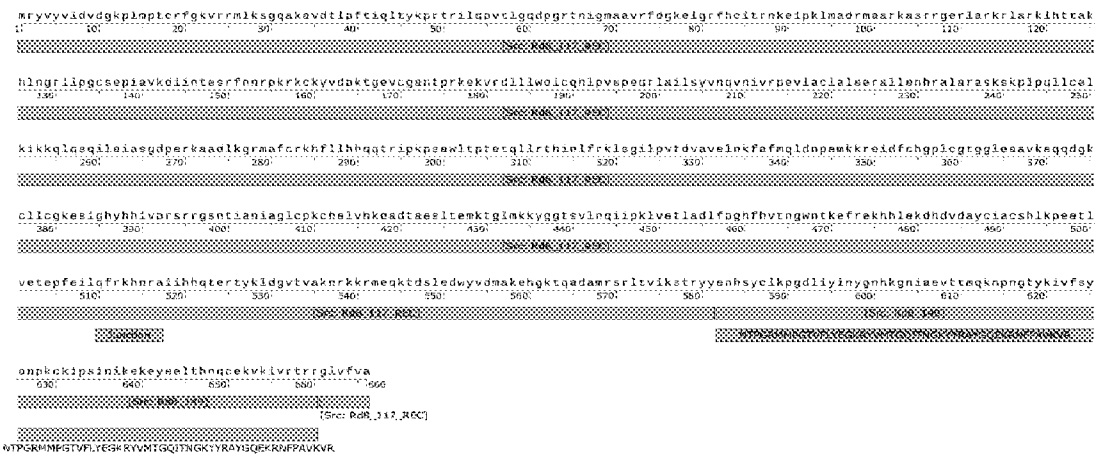


FIG. 93

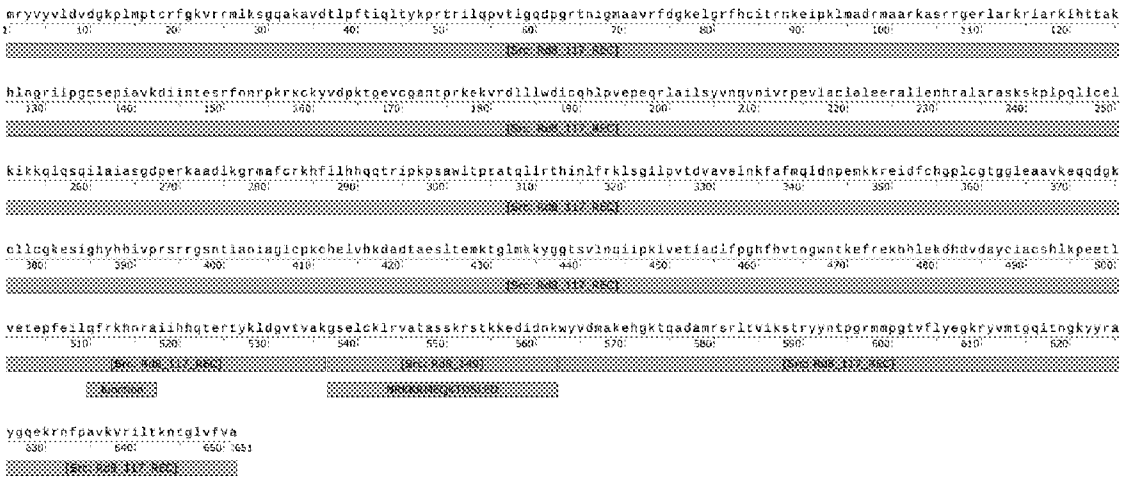

mryvyvldvdgkplmptcrfgkvrrmlksggakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklmadrmaarkasrrgerlarkriarkihttak
hlngriipgcsepiavkdiintesrfonrpkrkckyvdpktgevcgantprkekvrdlllwdicghlpvepeqrlailsyvngvnivrpevlaclaleeralleohralaraskskplpqlicel
kikkqlqsqilaiasgdperkaadlkgrmafcrkhfilhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
cllcgkesighyhhlvprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetiadlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetl
vetepfeilqfrkhnraiihhqtertykldgvtvakgselcklrvatasskrstkkedidnkwyvdmakehgktqadamrsrltvikstryyntpgrmmpgtvflyegkryvmtqqitngkyyra
ygqekrnfpavkvriltknrglvfva

FIG. 94

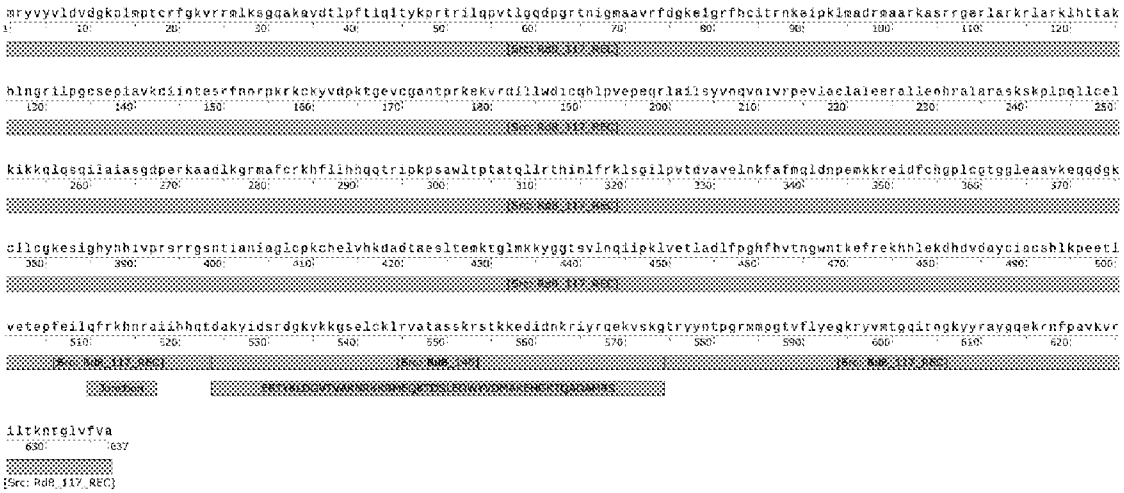

mryvyvldvdgkplmptcrfgkvrrmlksggakavdtlpftiqltykprtrilqpvtlgqdpgrtnigmaavrfdgkelgrfhcitrnkeipklmadrmaarkasrrgerlarkriarkihttak
hlngriipgcsepiavkdiintesrfonrpkrkckyvdpktgevcgantprkekvrdlllwdicghlpvepeqrlailsyvngvnivrpevlaclaleeralleohralaraskskplpqlicel
kikkqlqsqilaiasgdperkaadlkgrmafcrkhfilhhqqtripkpsawltptatqllrthinlfrklsgilpvtdvavelnkfafmqldnpemkkreidfchgplcgtggleaavkeqqdgk
cllcgkesighyhhlvprsrrgsntianiaglcpkchelvhkdadtaesltemktglmkkyggtsvlnqiipklvetiadlfpghfhvtngwntkefrekhhlekdhdvdayciacshlkpeetl
vetepfeilqfrkhnraiihhqtdakyidsrdgkvkkgselcklrvatasskrstkkedidnkriyrqekvskgtryyntpgrmmpgtvflyegkryvmtqqitngkyyrayqqekrnfpavkvr
iltknrglvfva

FIG. 95

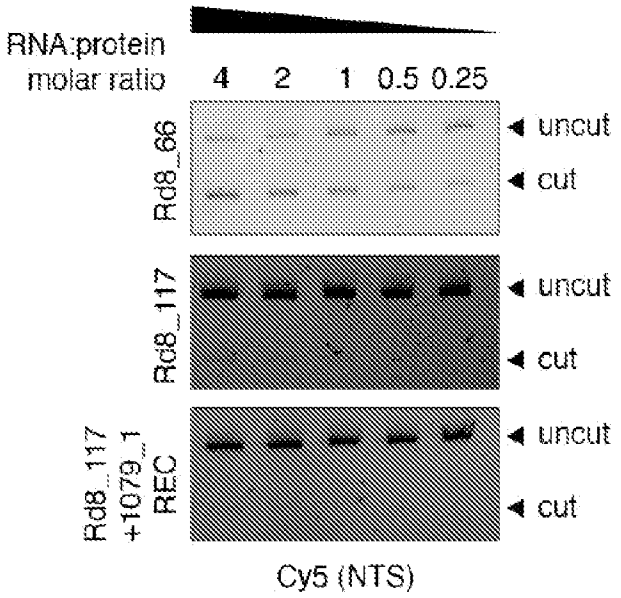


FIG. 96A

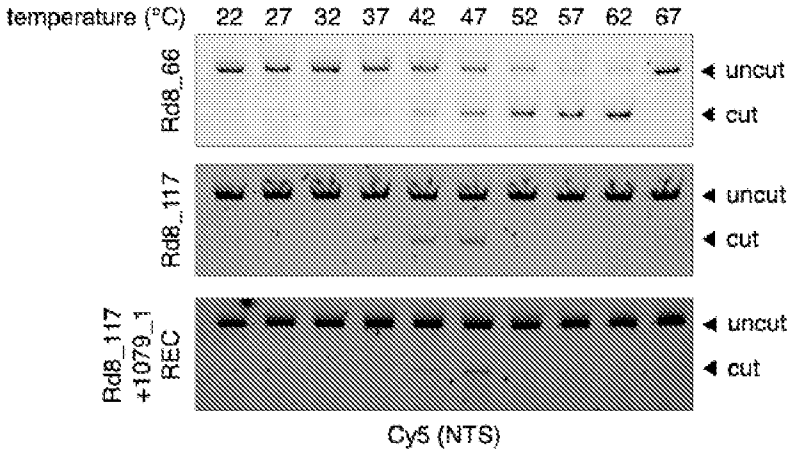


FIG. 96B

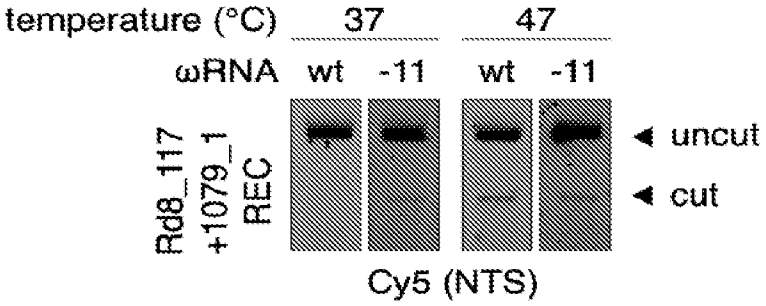


FIG. 96C

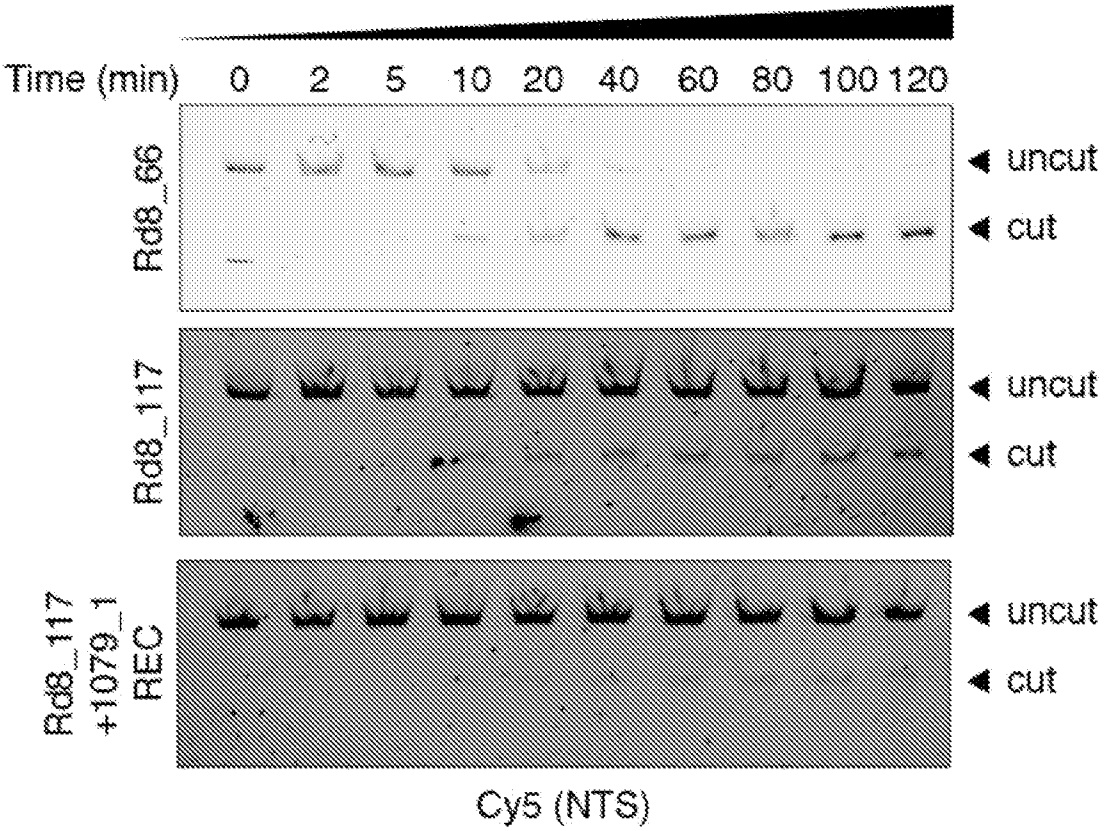

Time (min)
0
2
5
10
20
40
60
80
100
120
Rd8_66
uncut
cut
Rd8_117
uncut
cut
Rd8_117 +1079_1 REC
uncut
cut
Cy5 (NTS)

FIG. 96D

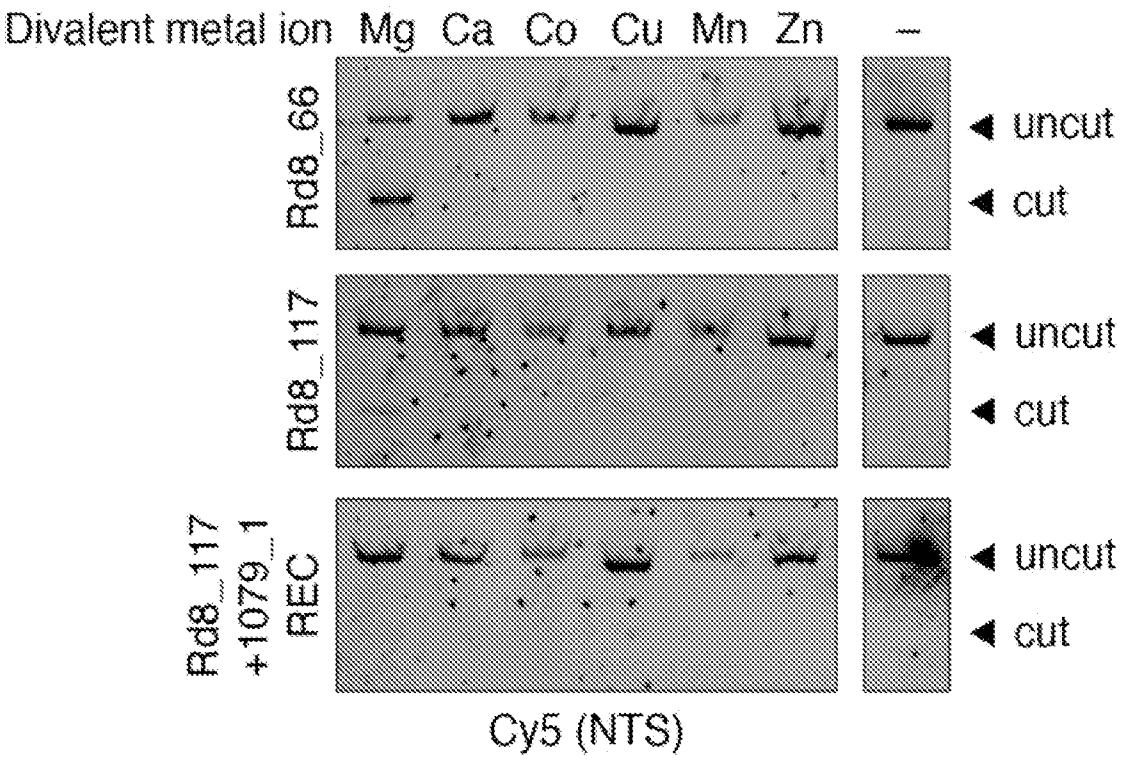


FIG. 96E

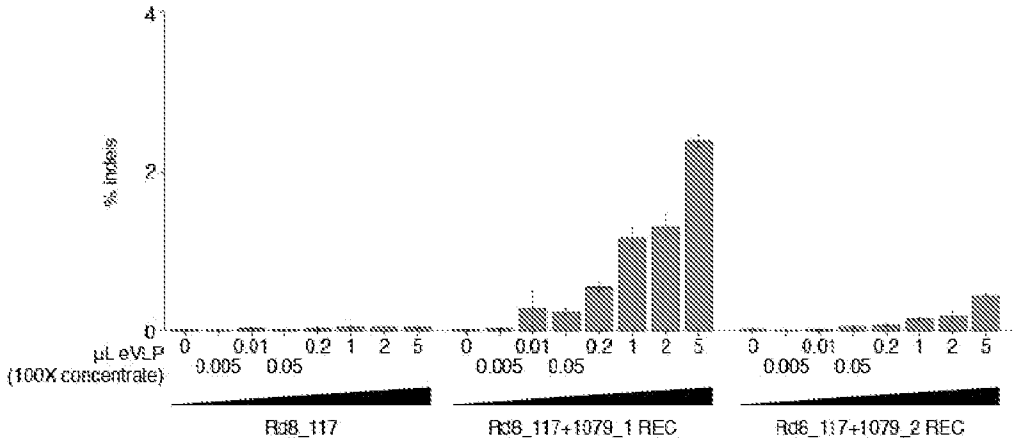


FIG. 97

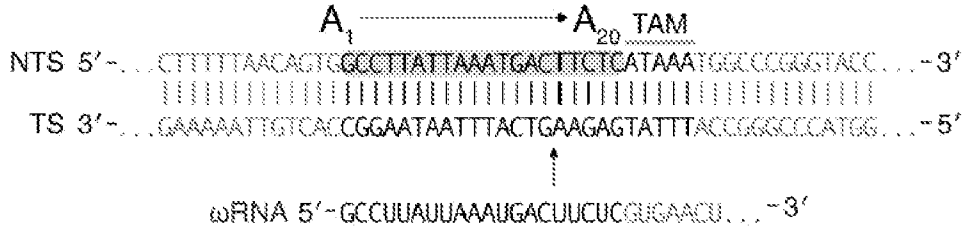


FIG. 98A

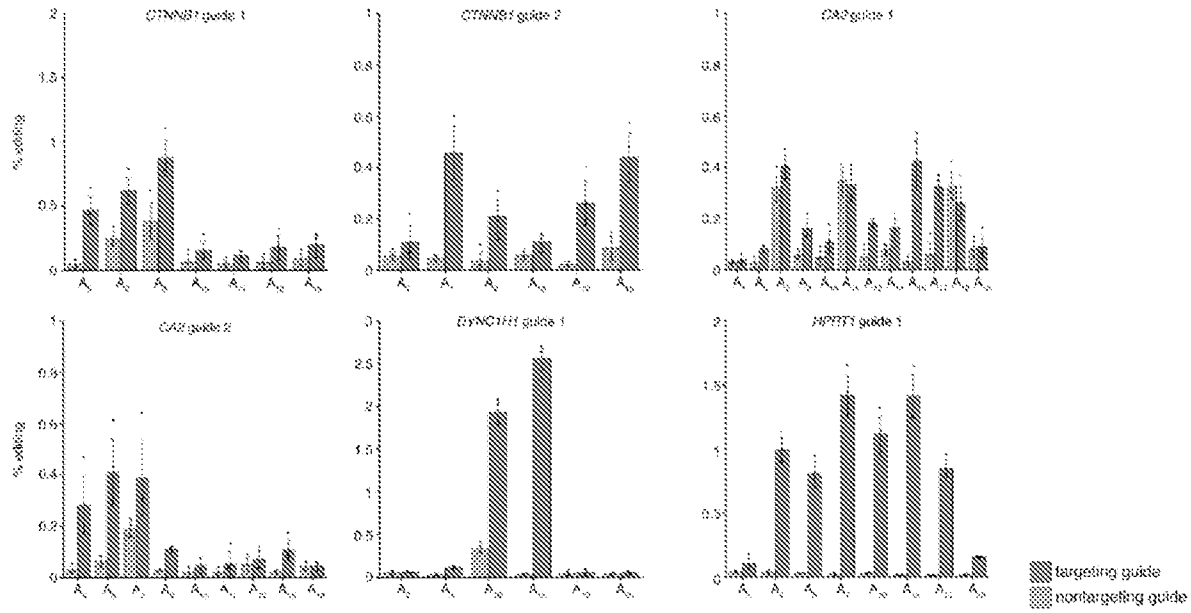


FIG. 98B

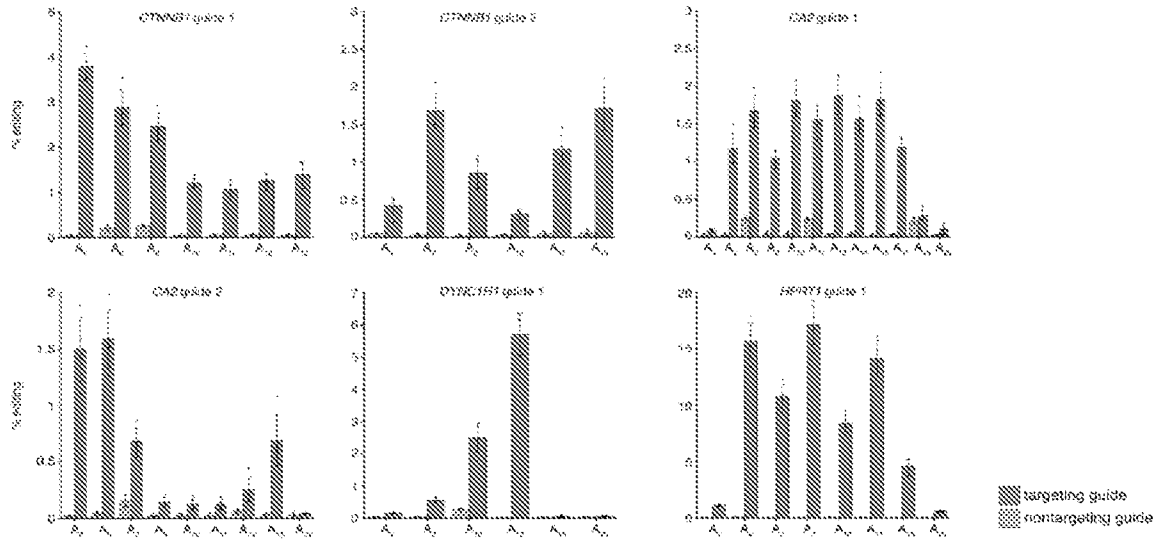


FIG. 98C

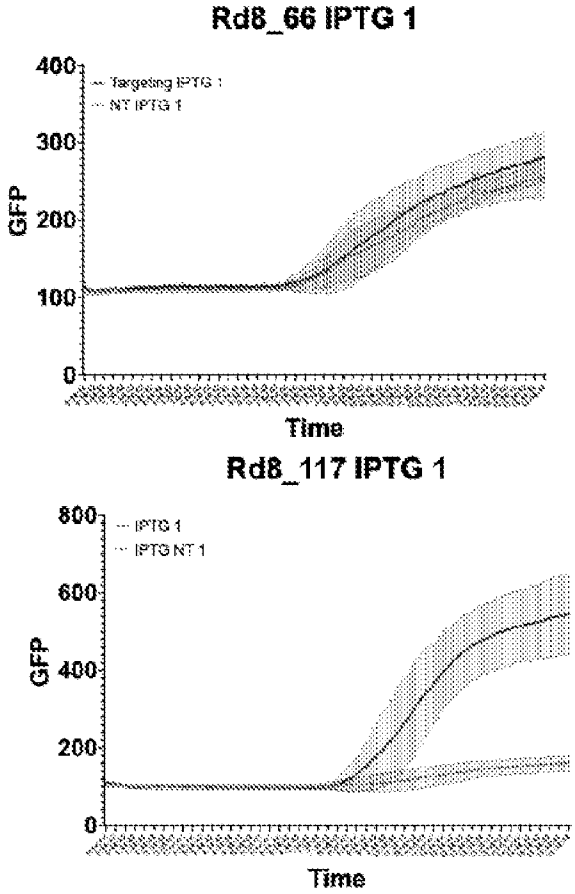

Rd8_66 IPTG 1
Targeting IPTG 1
NT IPTG 1
400
300
200
100
0
GFP
Time
Rd8_117 IPTG 1
IPTG 1
IPTG NT 1
800
600
400
200
0
GFP
Time

FIG. 99A

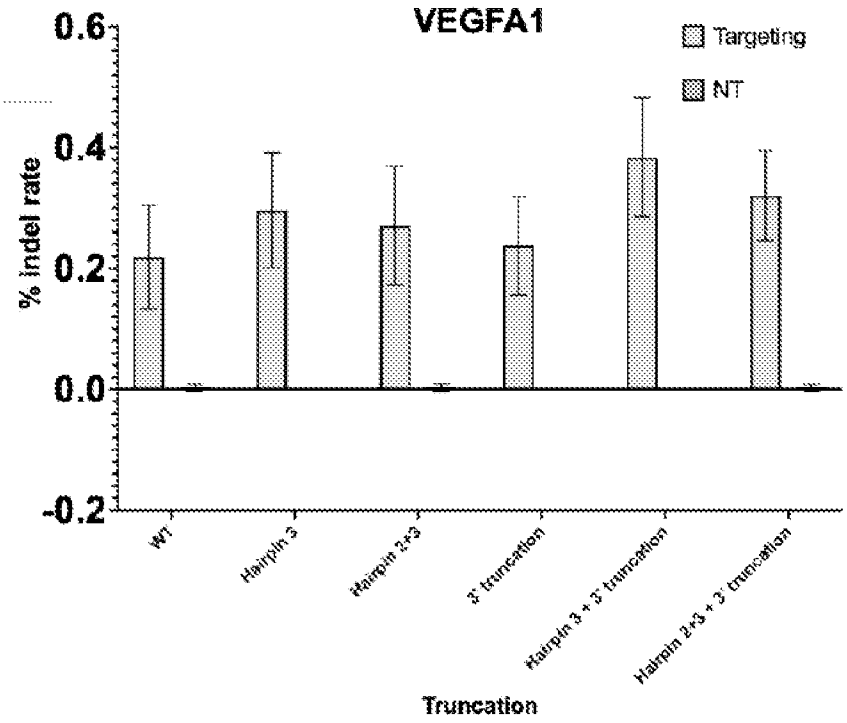


FIG. 99B

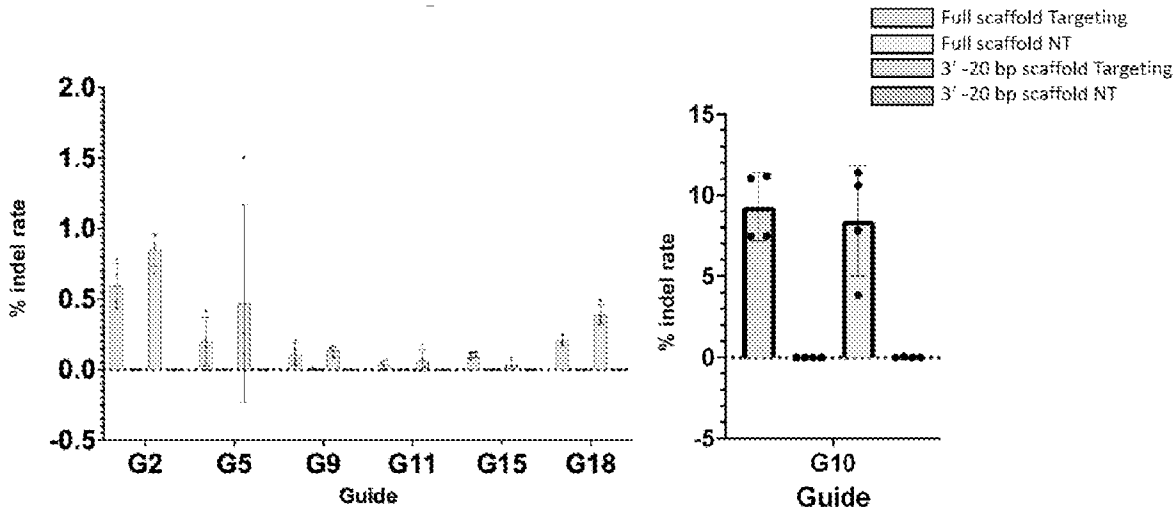


FIG. 100

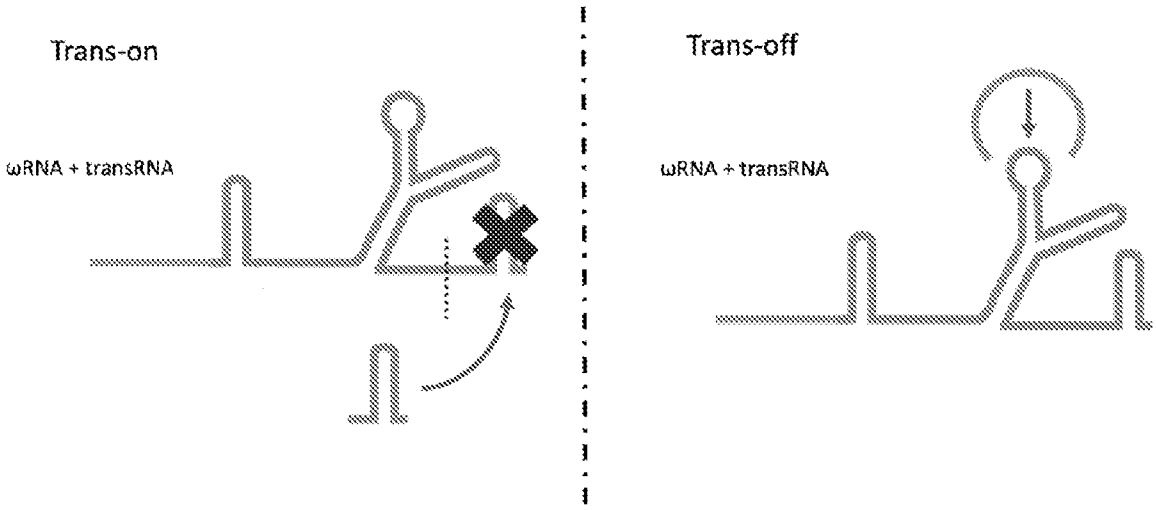


FIG. 101

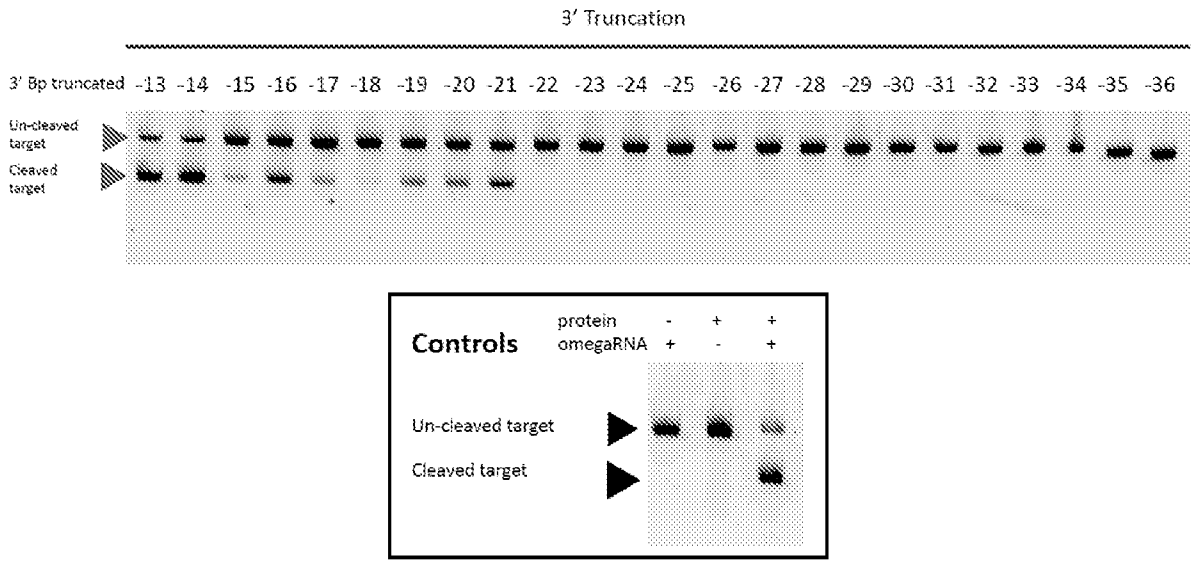


FIG. 102A

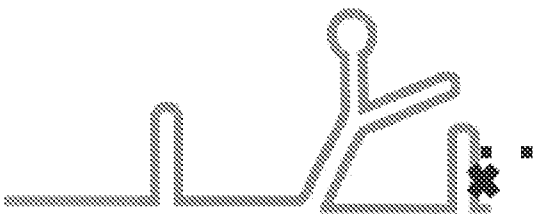


FIG. 102B

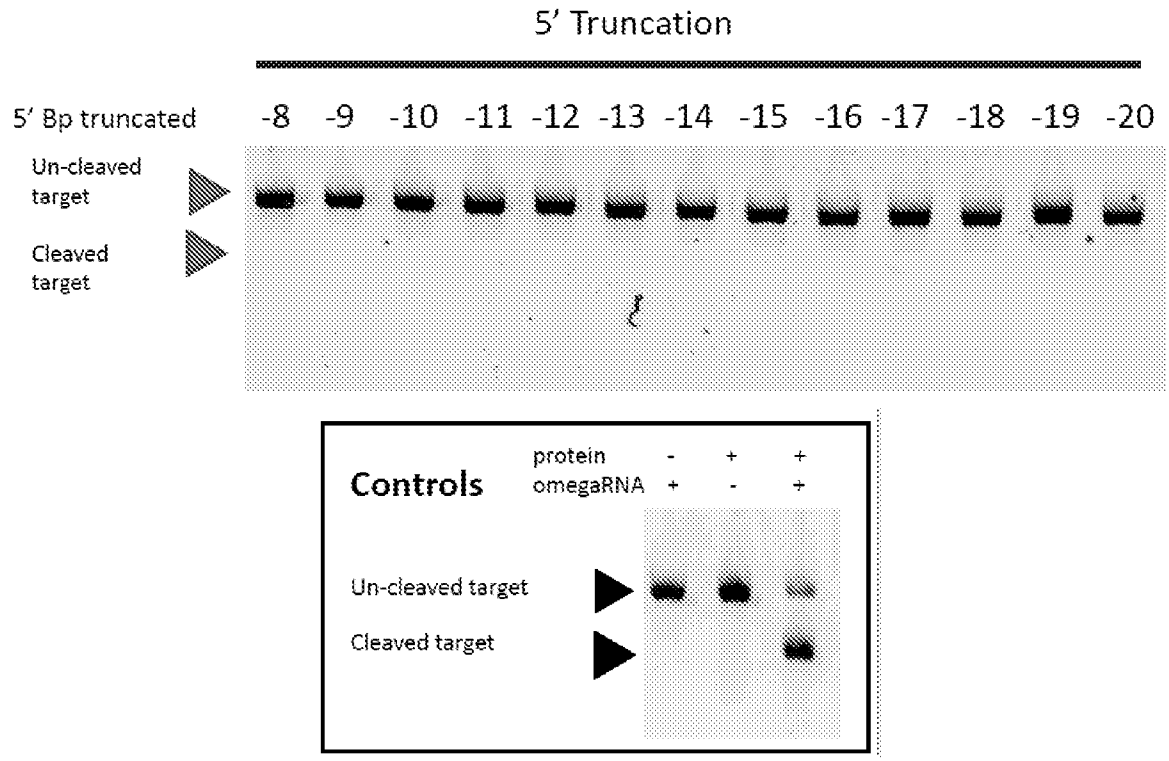


FIG. 102C

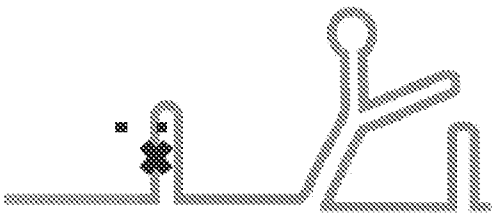
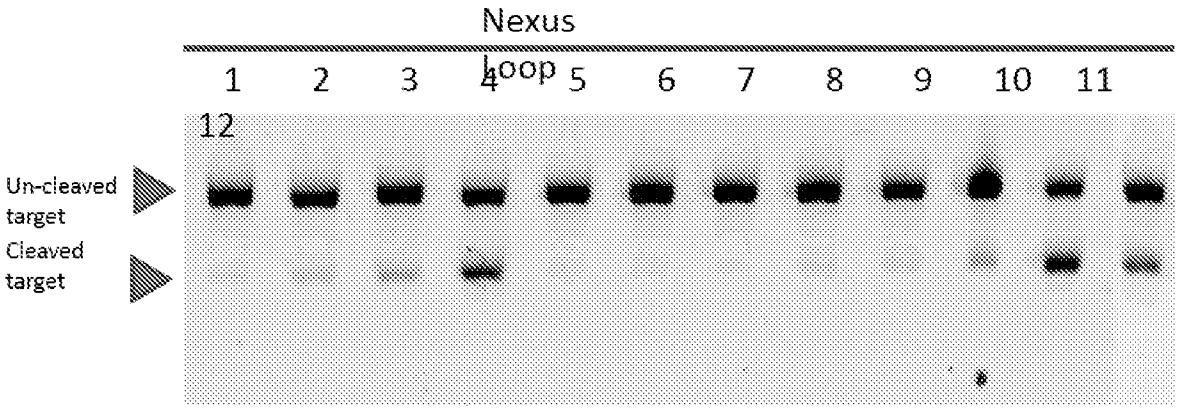


FIG. 102D



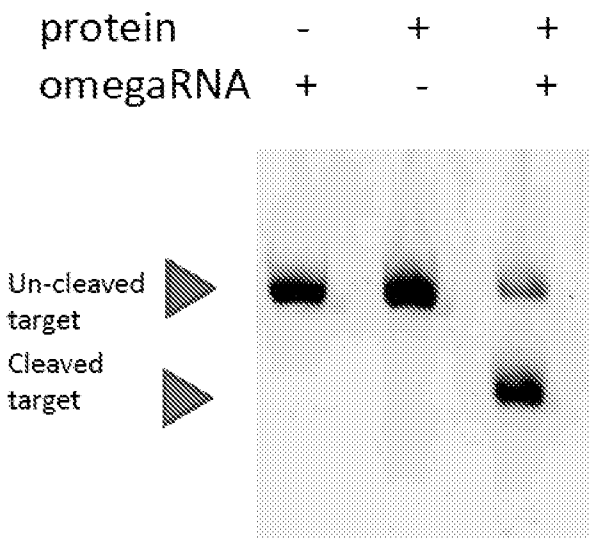


FIG. 103A

Original	GATTCCAA
Nexus Loop 1	GATTTCCAA
Nexus Loop 2	GAGTTTCCAA
Nexus Loop 3	GAAGTTTCCAA
Nexus Loop 4	GATTCCATA
Nexus Loop 5	GATTGCCAA with reverse complement in nexus
Nexus Loop 6	GATTCCCAA with reverse complement in nexus
Nexus Loop 7	GATTGCCCAA with reverse complement in nexus
Nexus Loop 8	GATTCGCCAA with reverse complement in nexus
Nexus Loop 9	GATTGCCAA
Nexus Loop 10	GATTCCCAA
Nexus Loop 11	GATTGCCCAA
Nexus Loop 12	GATTCGCCAA

FIG. 103B

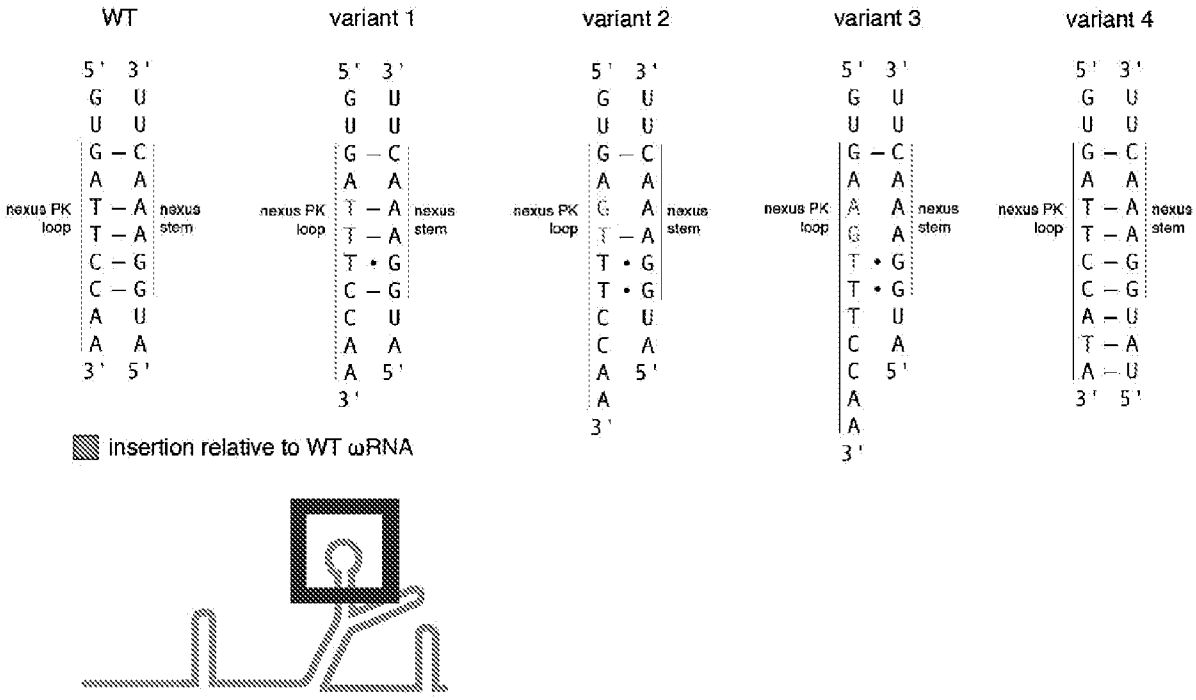
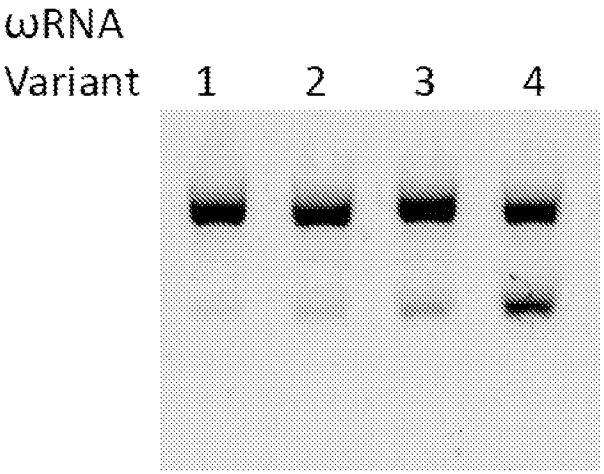

WT
variant 1
variant 2
variant 3
variant 4
nexus PK loop
nexus stem
insertion relative to WT ωRNA

FIG. 104A



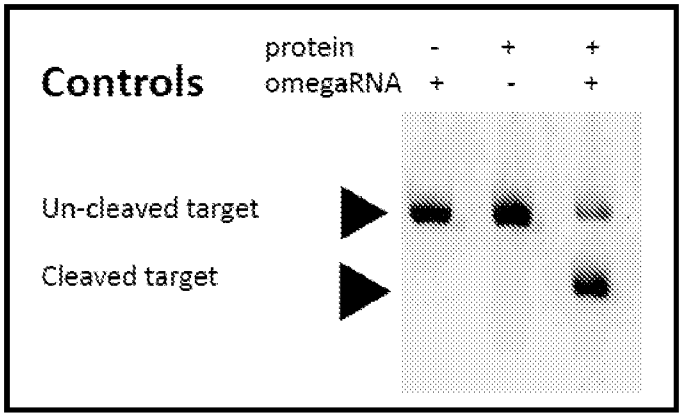


FIG. 104B

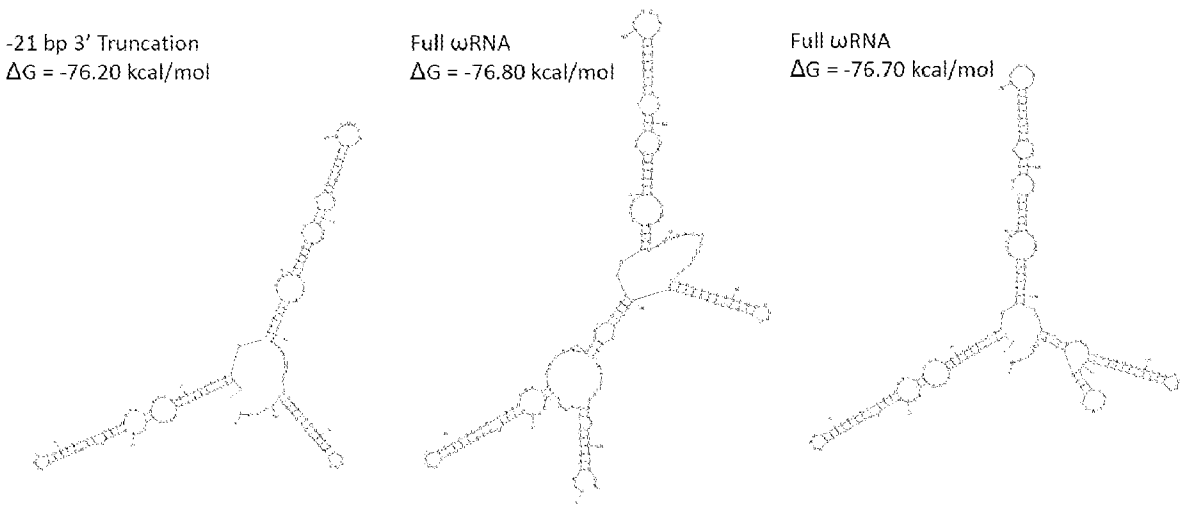


FIG. 105

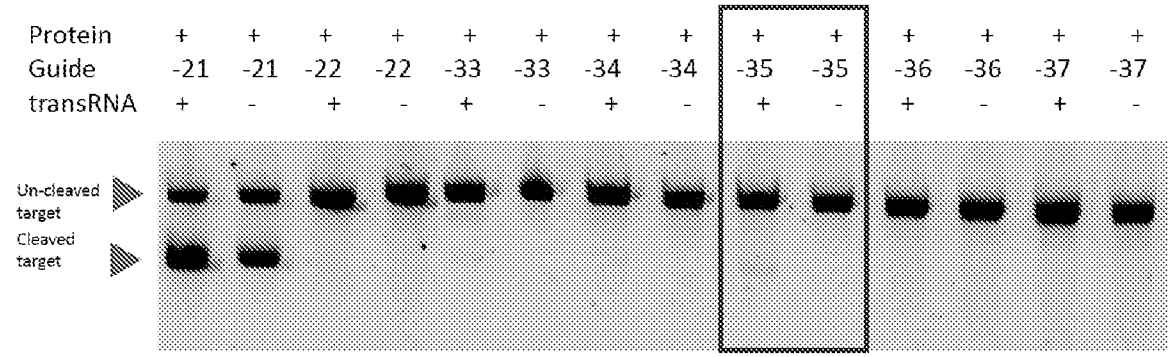


FIG. 106A

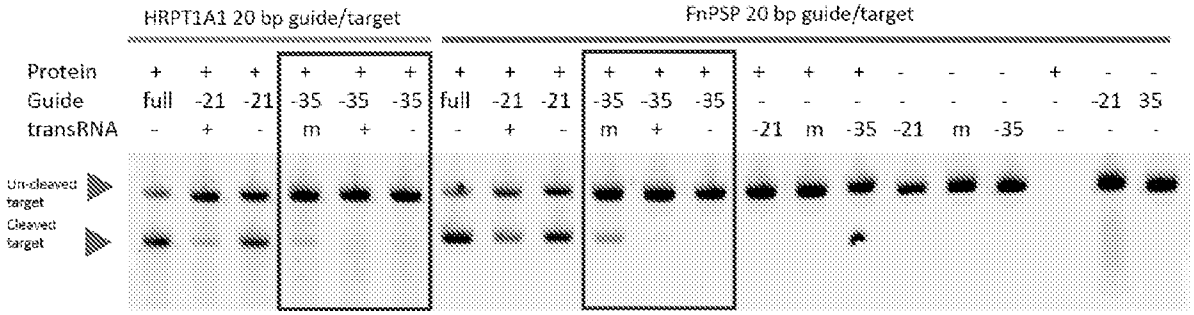


FIG. 106B

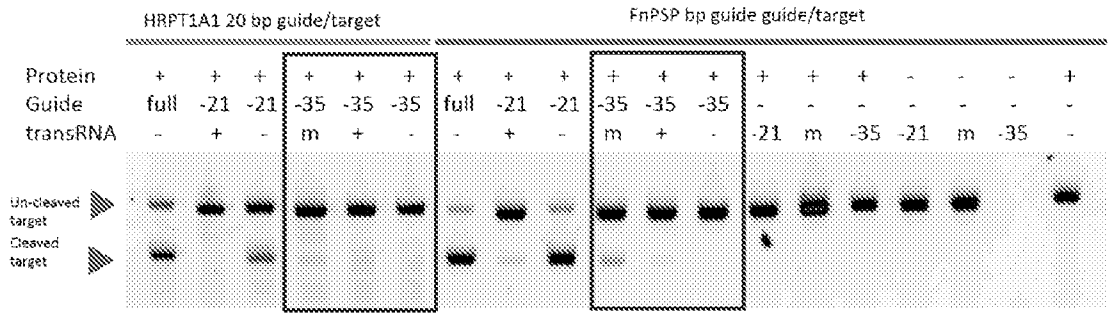


FIG. 106C

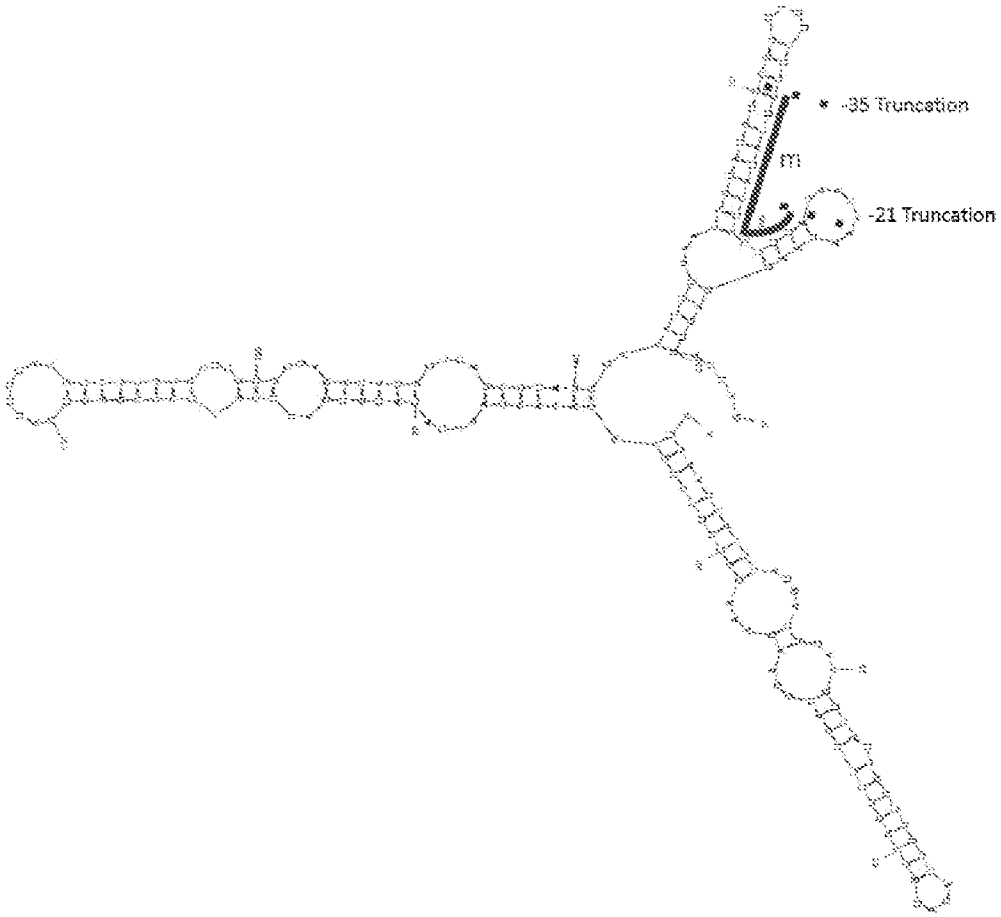


FIG. 106D

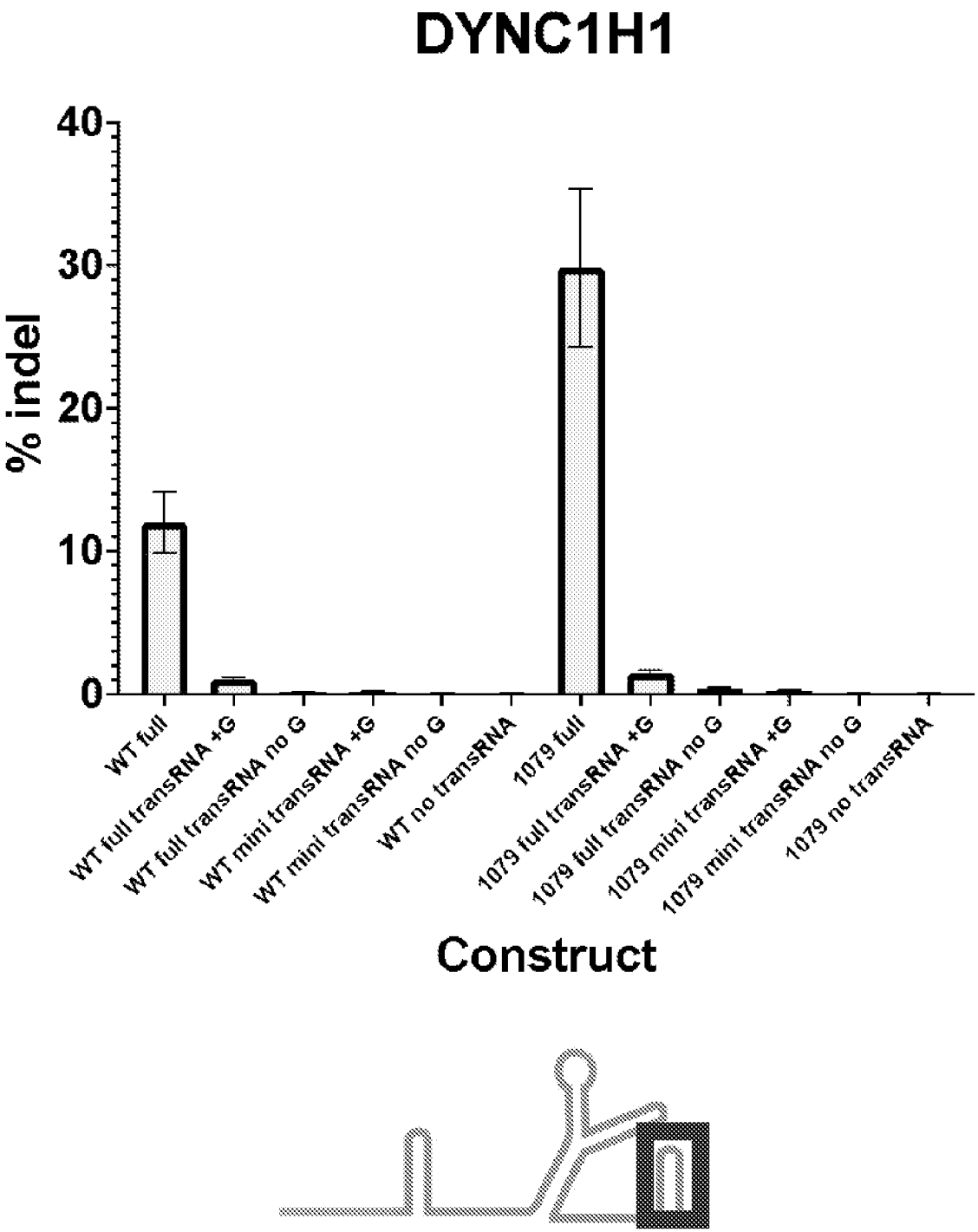

DYNC1H1
% indel
40
30
20
10
0
WT full
WT full transRNA +G
WT full transRNA no G
WT mini transRNA +G
WT mini transRNA no G
WT no transRNA
1079 full
1079 full transRNA +G
1079 full transRNA no G
1079 mini transRNA +G
1079 mini transRNA no G
1079 no transRNA
Construct

FIG. 107A

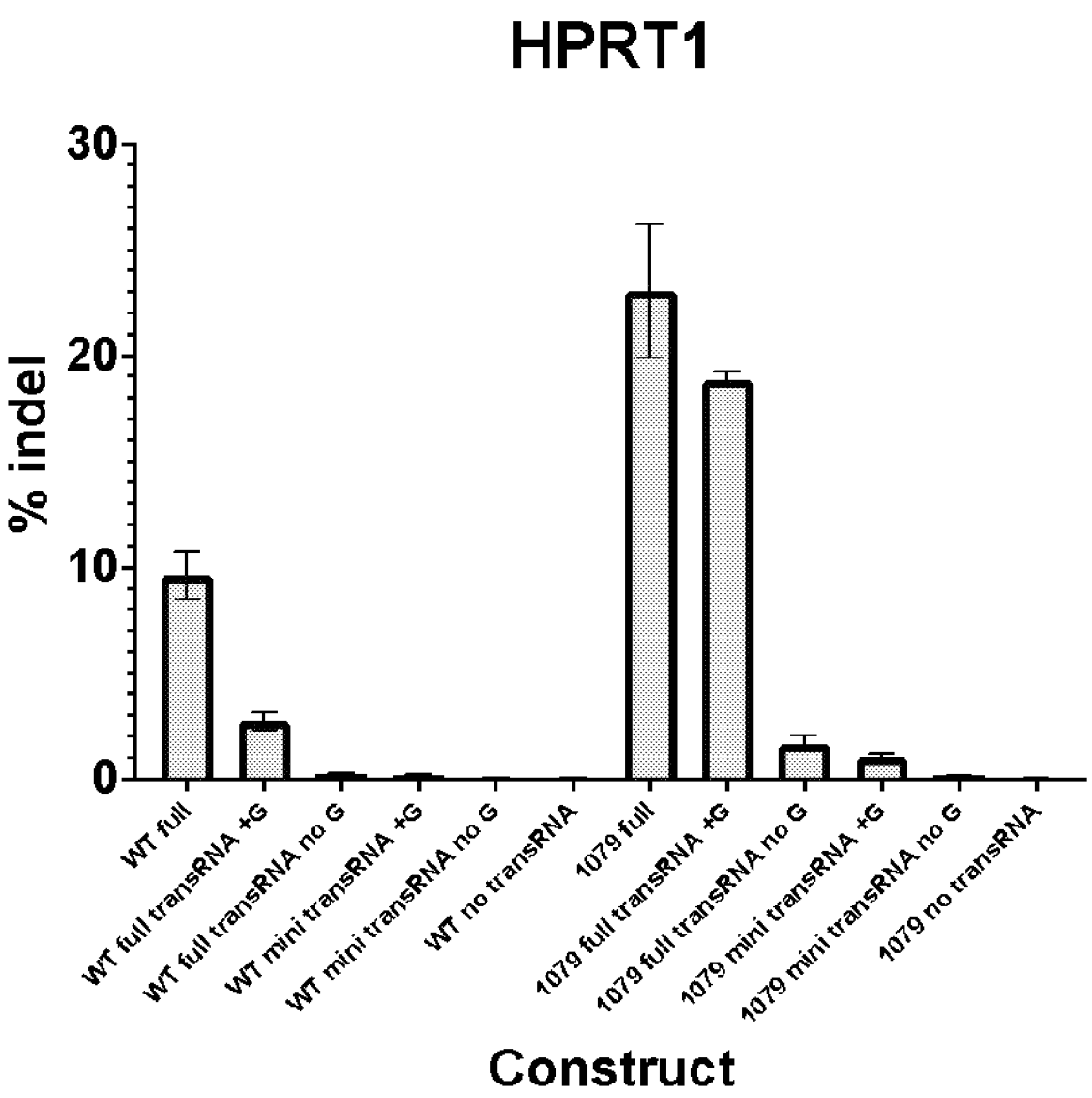

HPRT1
% indel
30
20
10
0
WT full
WT full transRNA +G
WT full transRNA no G
WT mini transRNA +G
WT mini transRNA no G
WT no transRNA
1079 full
1079 full transRNA +G
1079 full transRNA no G
1079 mini transRNA +G
1079 mini transRNA no G
1079 no transRNA
Construct

FIG. 107B

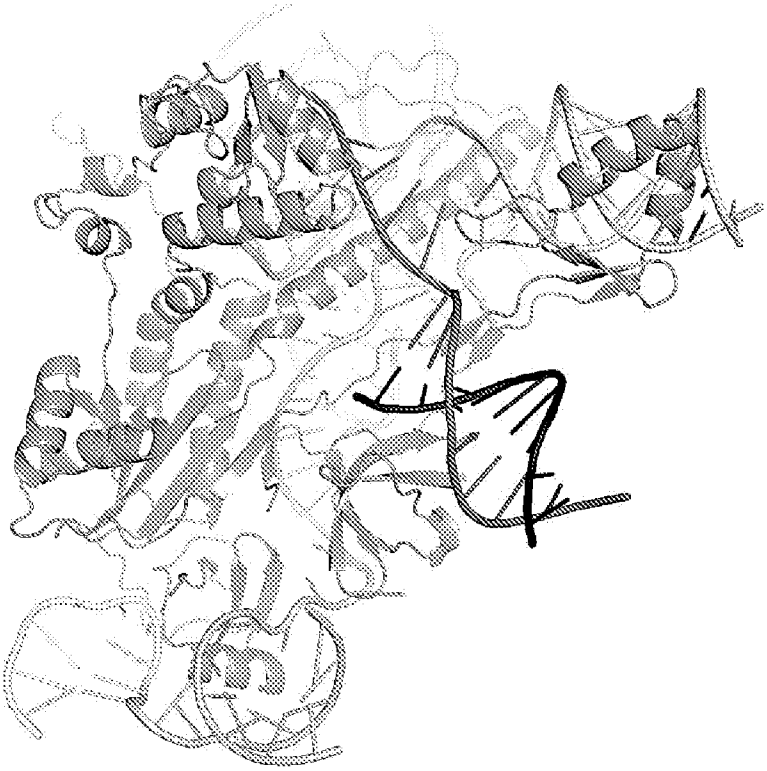


FIG. 108

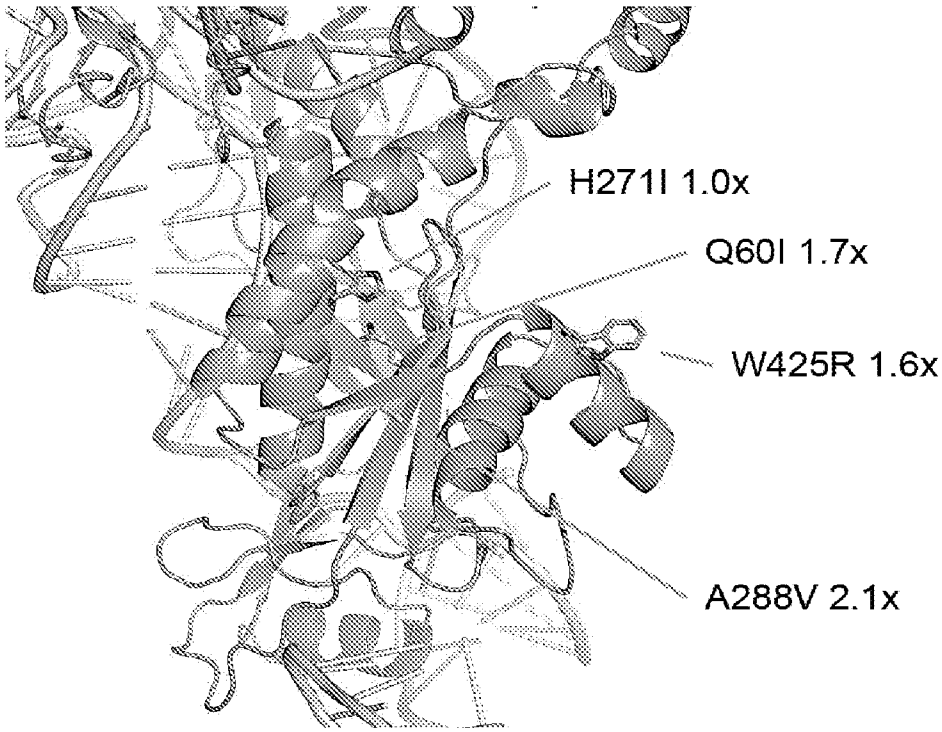


FIG. 109

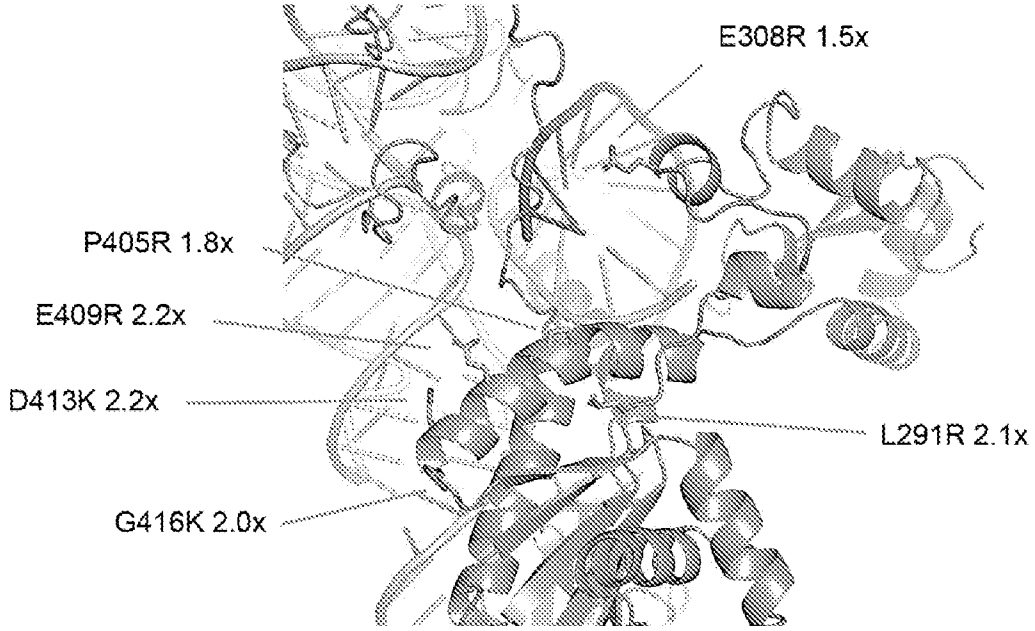


FIG. 110

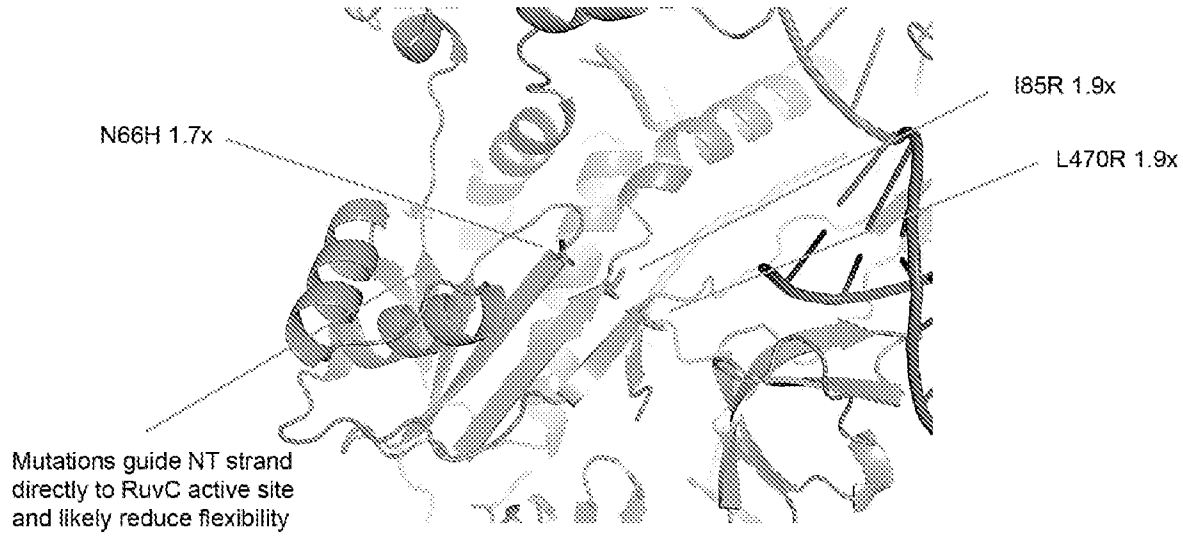


FIG. 111

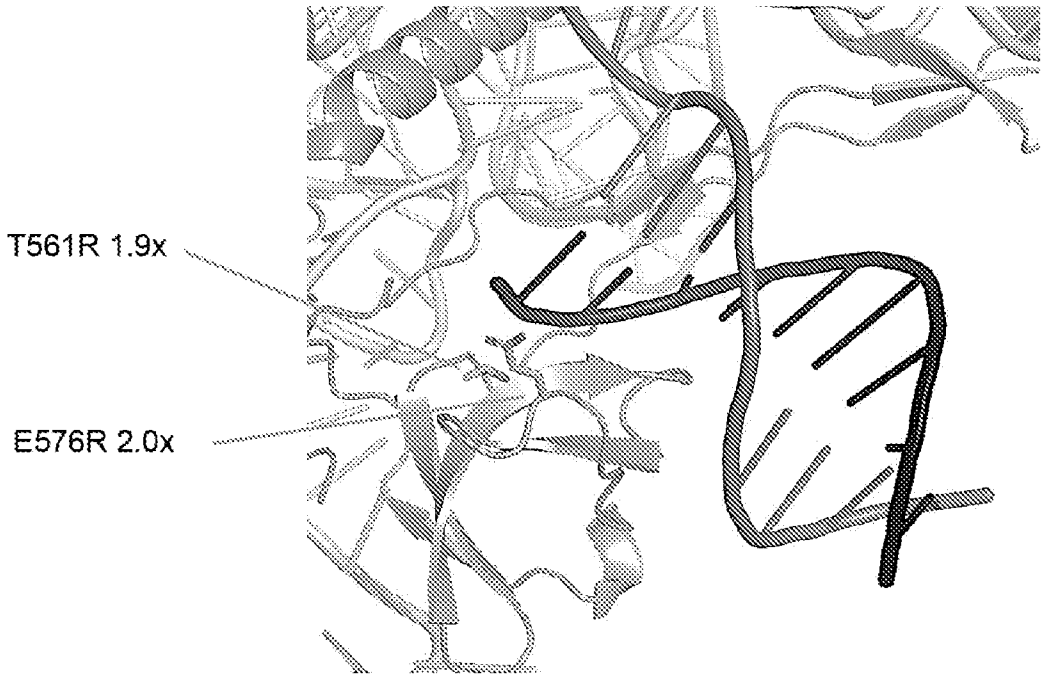


FIG. 112

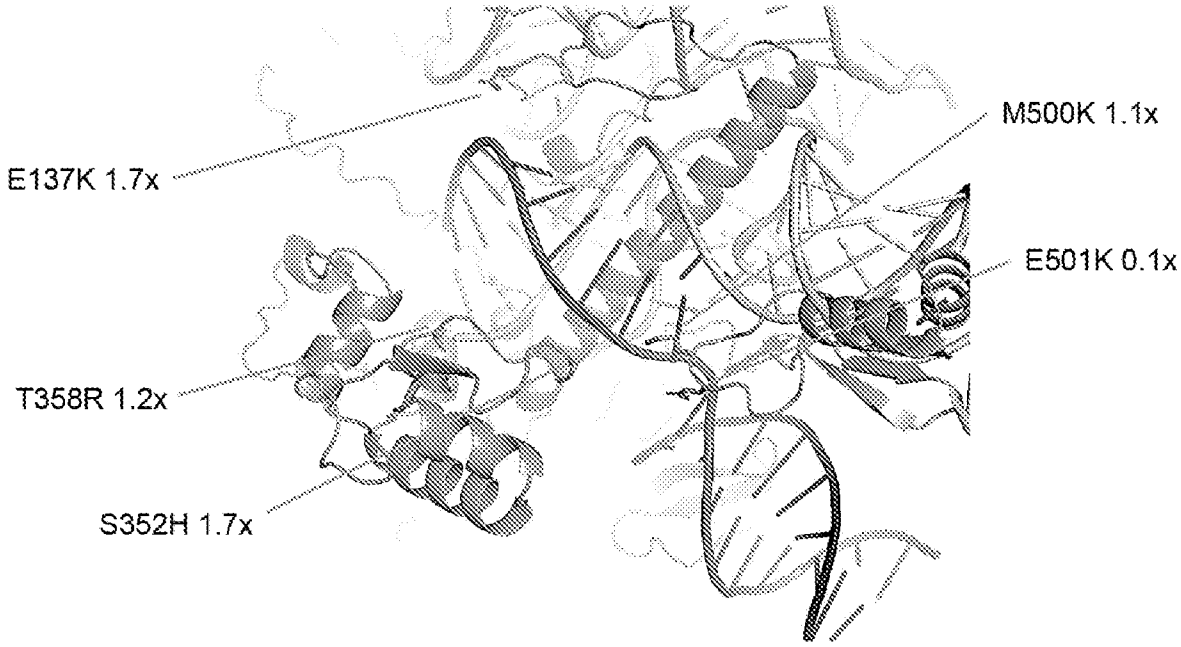


FIG. 113

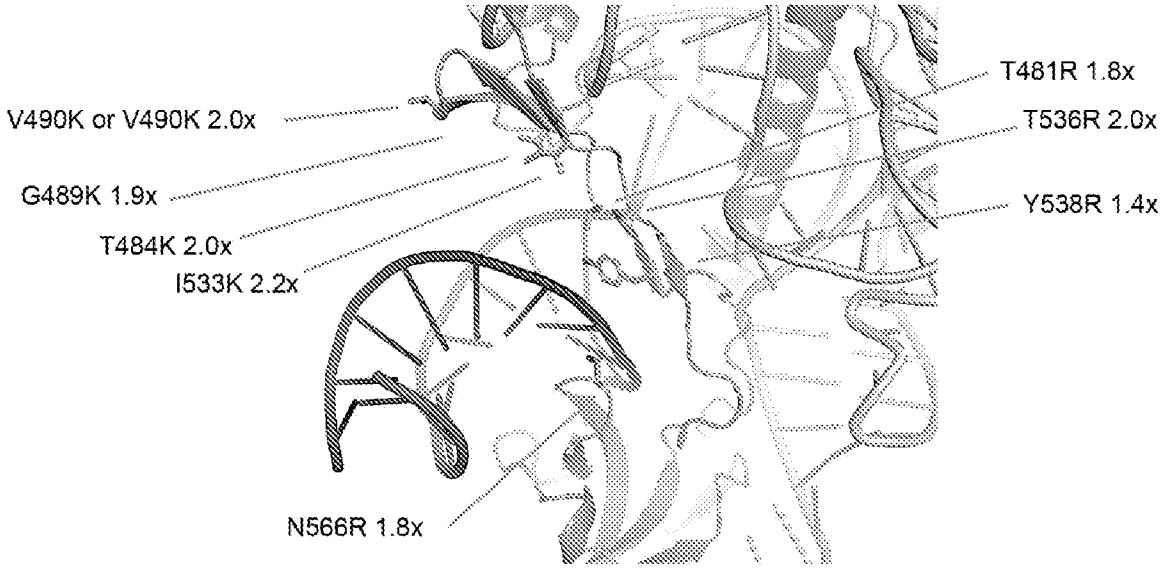


FIG. 114

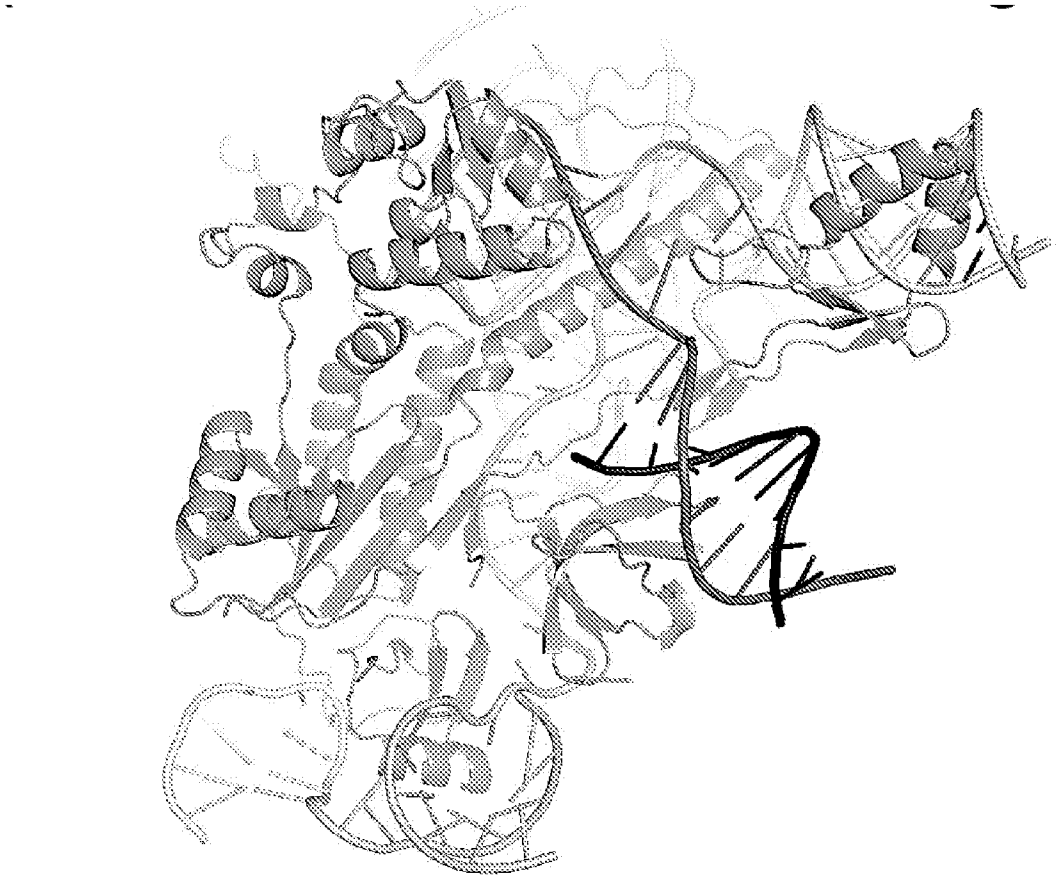


FIG. 115

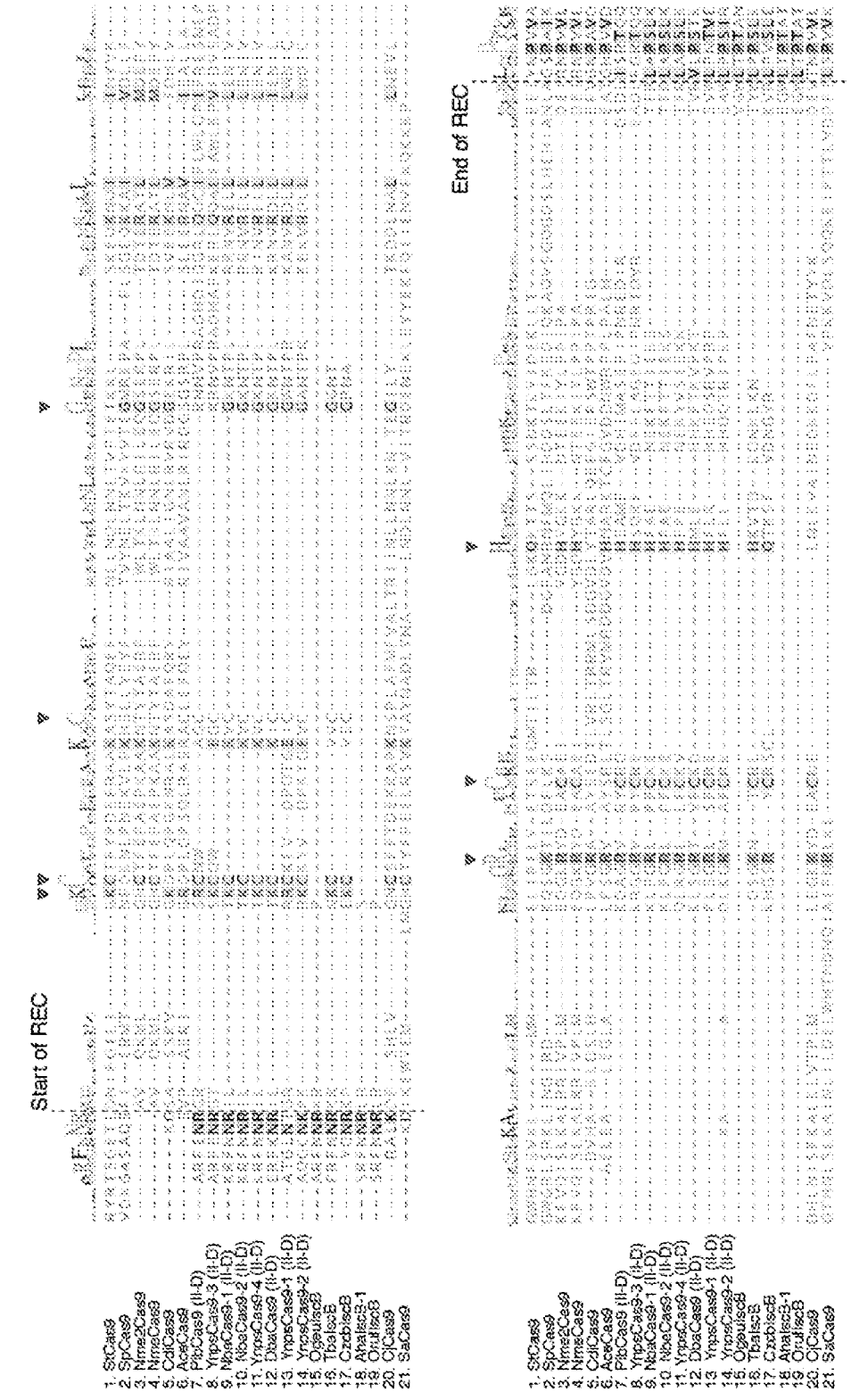


FIG. 116A-116B

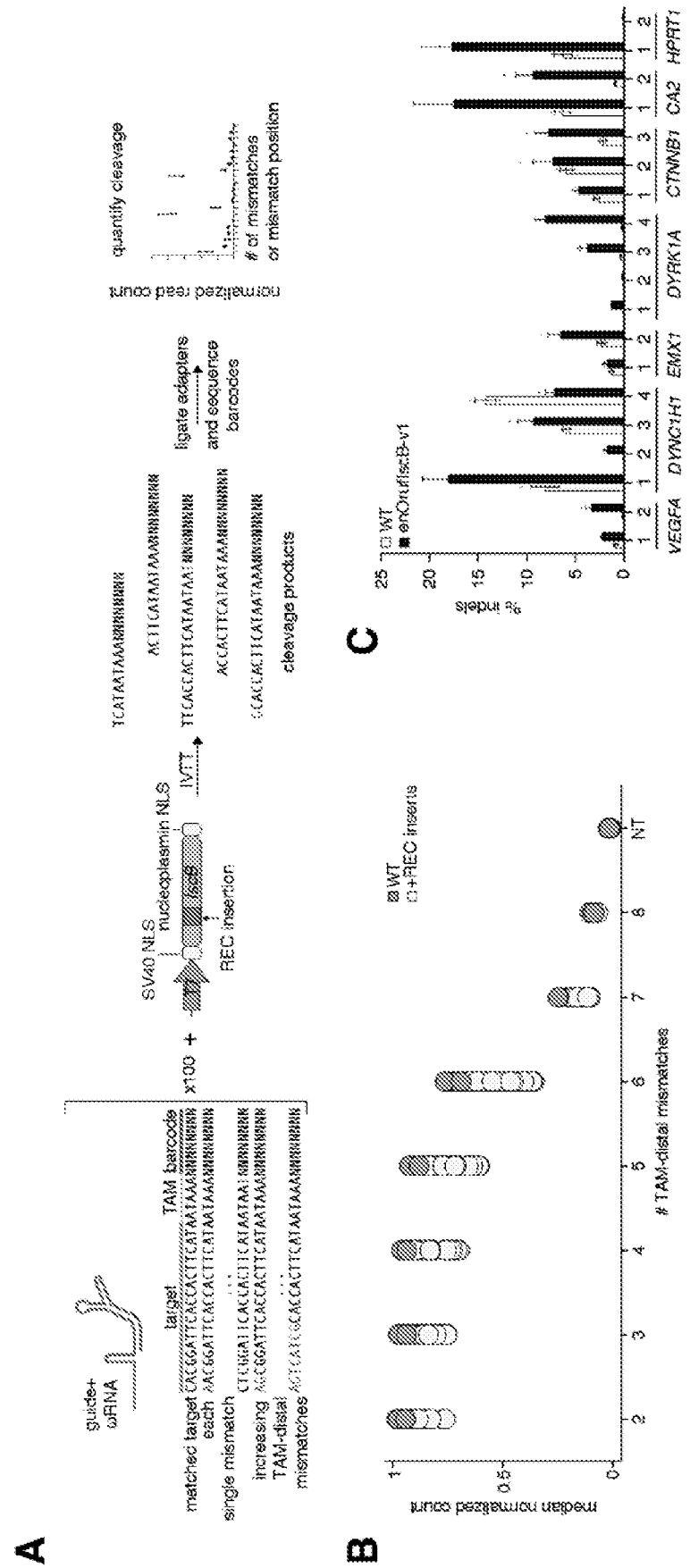


FIG. 117A-117C

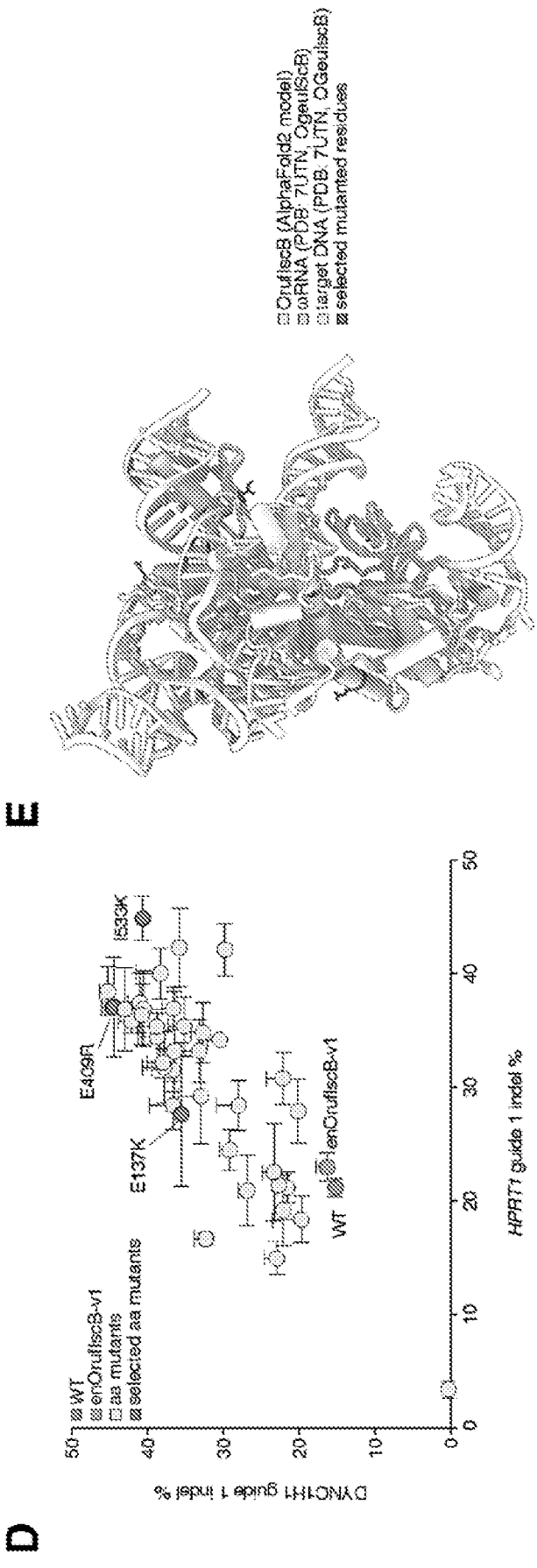


FIG. 117D-117E

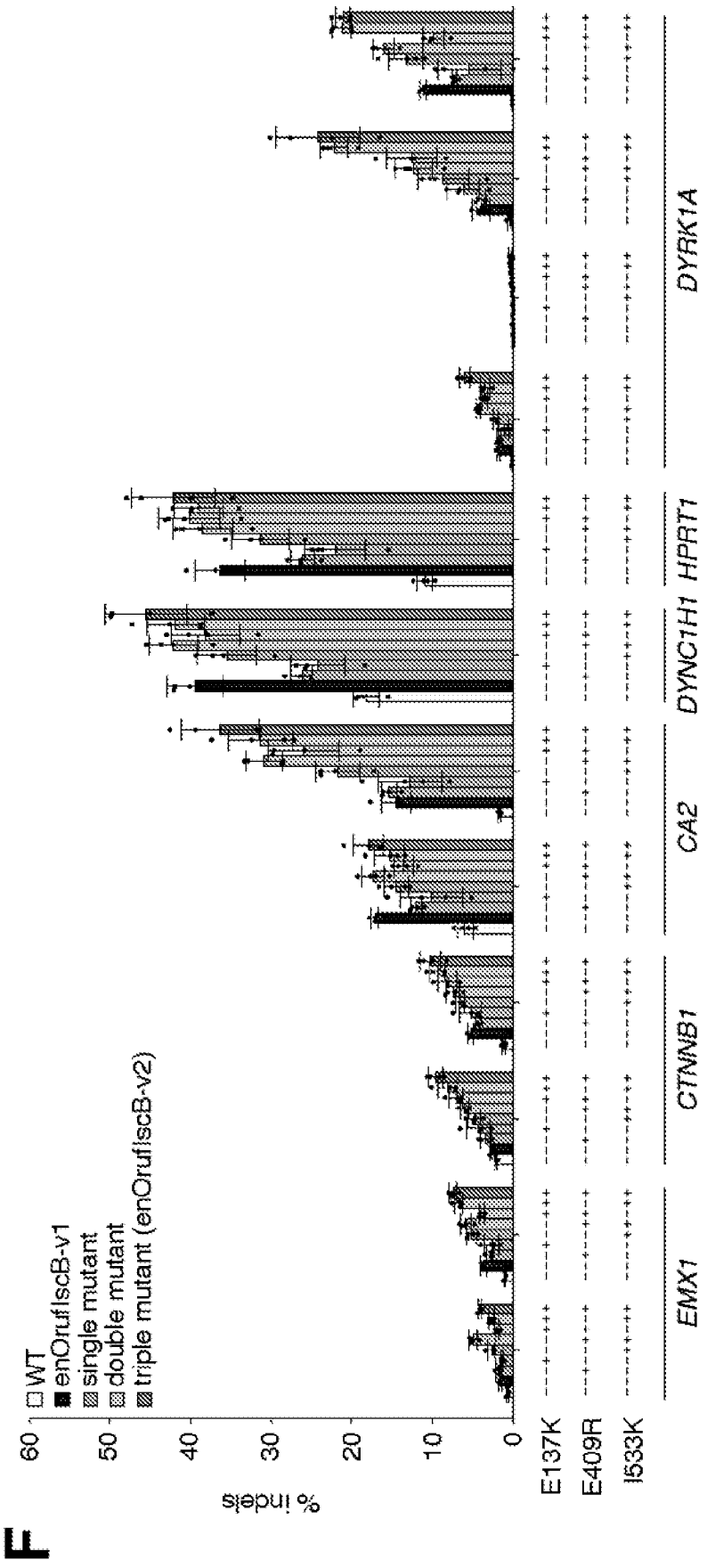

F
% indels
60
50
40
30
20
10
0
WT
enOruflscB-v1
single mutant
double mutant
triple mutant (enOruflscB-v2)
E137K
E409R
I533K
EMX1
CTNNB1
CA2
DYNC1H1
HPRT1
DYRK1A

FIG. 117F

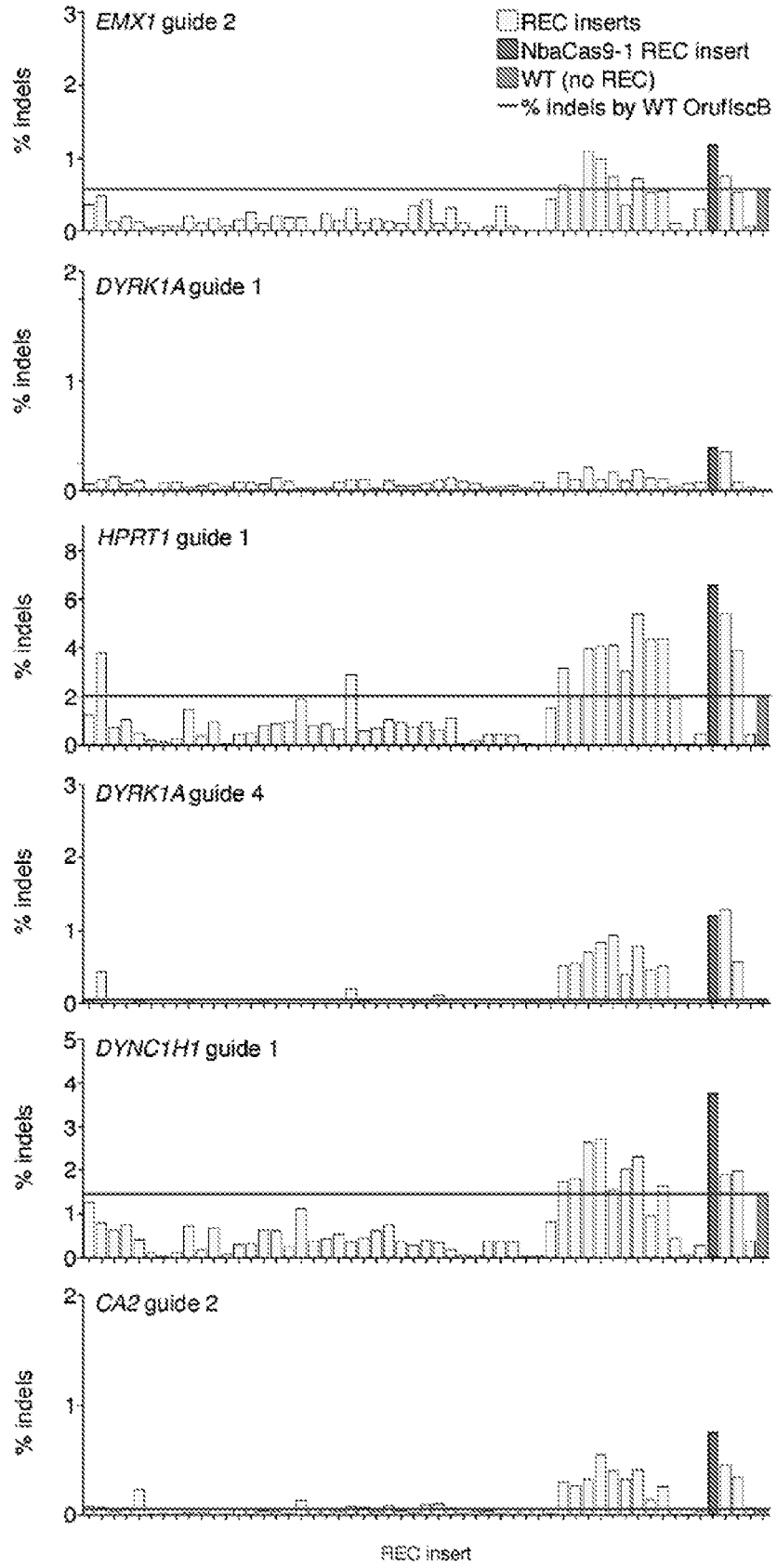

EMX1 guide 2
REC inserts
NbaCas9-1 REC insert
WT (no REC)
% indels by WT OruflscB
DYRK1A guide 1
HPRT1 guide 1
DYRK1A guide 4
DYNC1H1 guide 1
CA2 guide 2
% indels
REC insert

FIG. 118

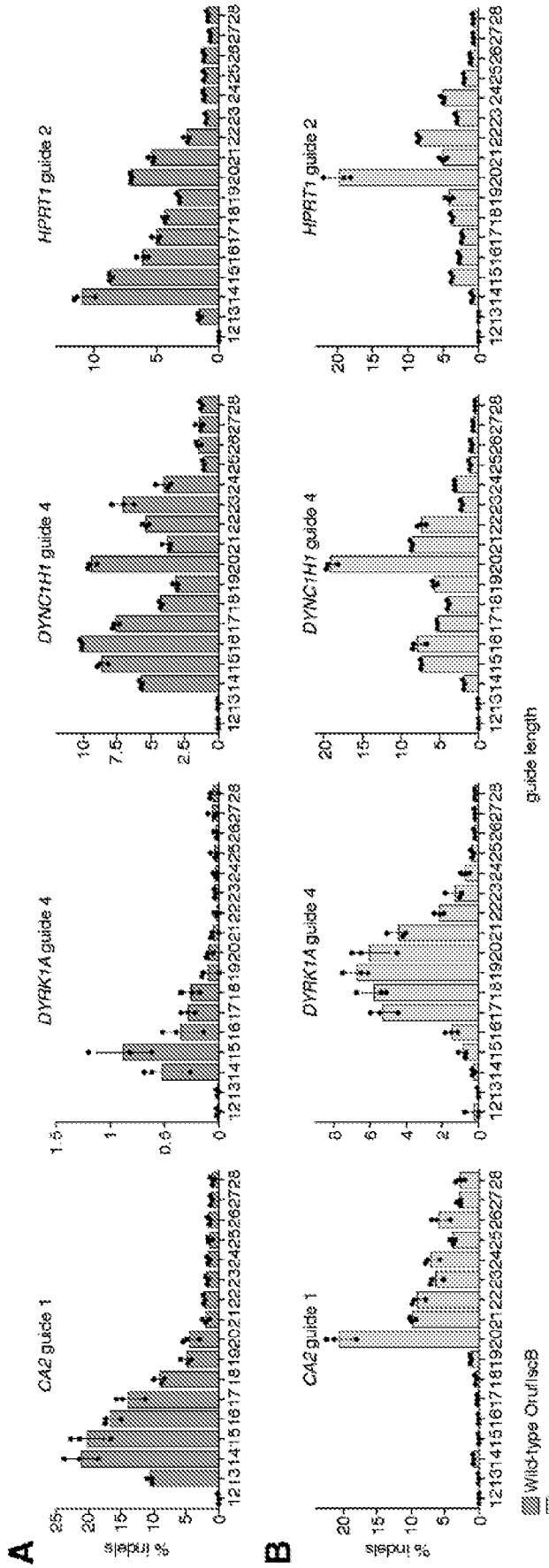


FIG. 119A-119B

A

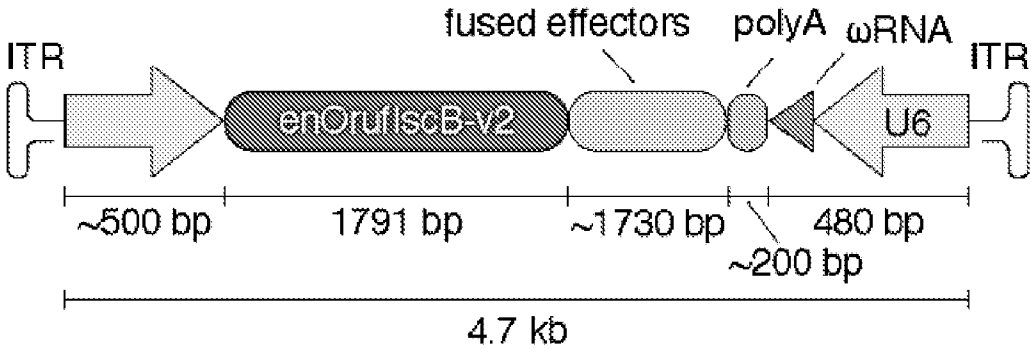

ITR
fused effectors
polyA
ωRNA
ITR
enOrufisoB-v2
U6
~500 bp
1791 bp
~1730 bp
480 bp
~200 bp
4.7 kb

FIG. 120A

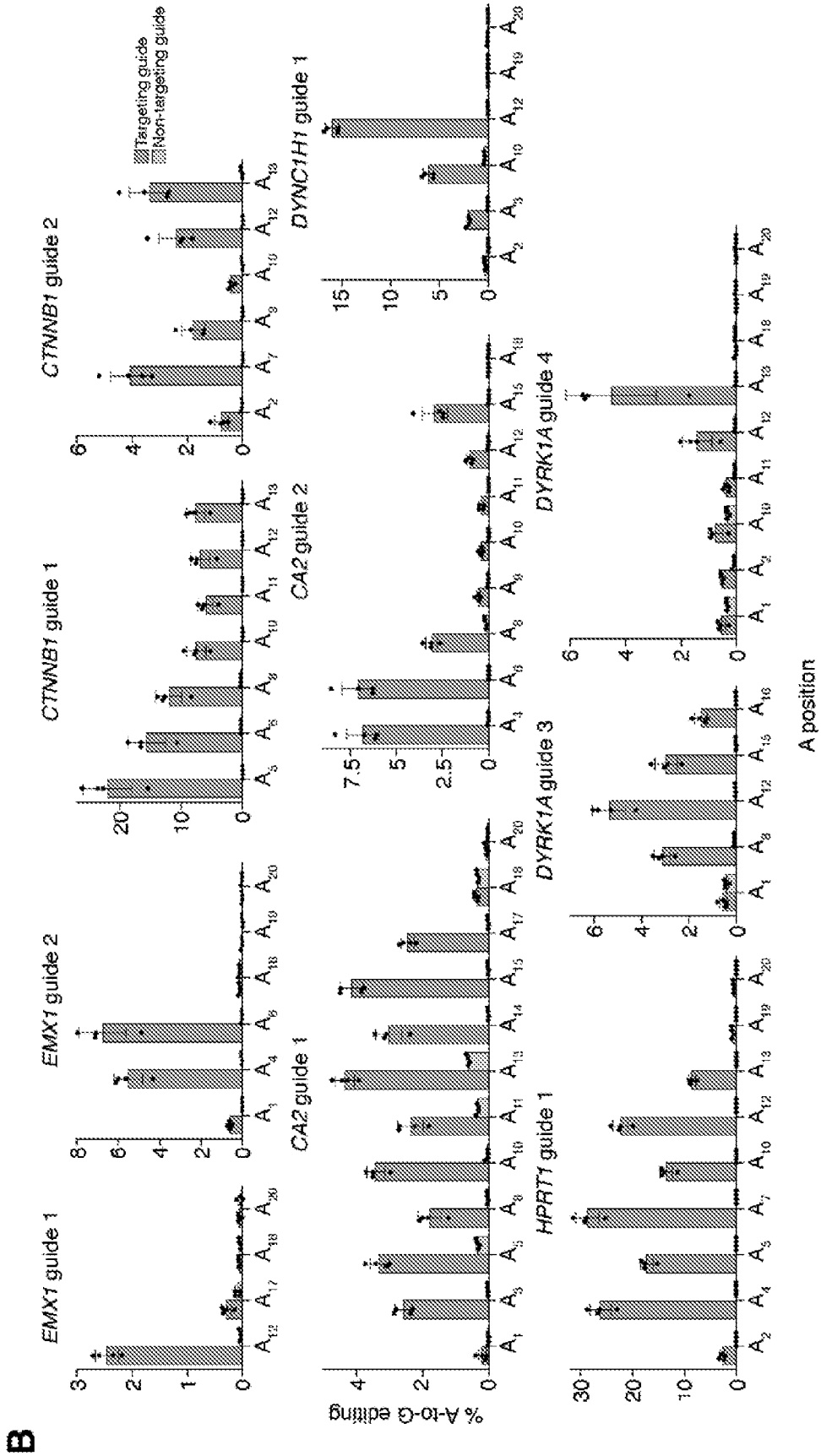

B
EMX1 guide 1
EMX1 guide 2
CTNNB1 guide 1
CTNNB1 guide 2
CA2 guide 1
CA2 guide 2
HPRT1 guide 1
DYRK1A guide 3
DYRK1A guide 4
DYNC1H1 guide 1
Targeting guide
Non-targeting guide
% A-to-G editing
A position

FIG. 120B

C

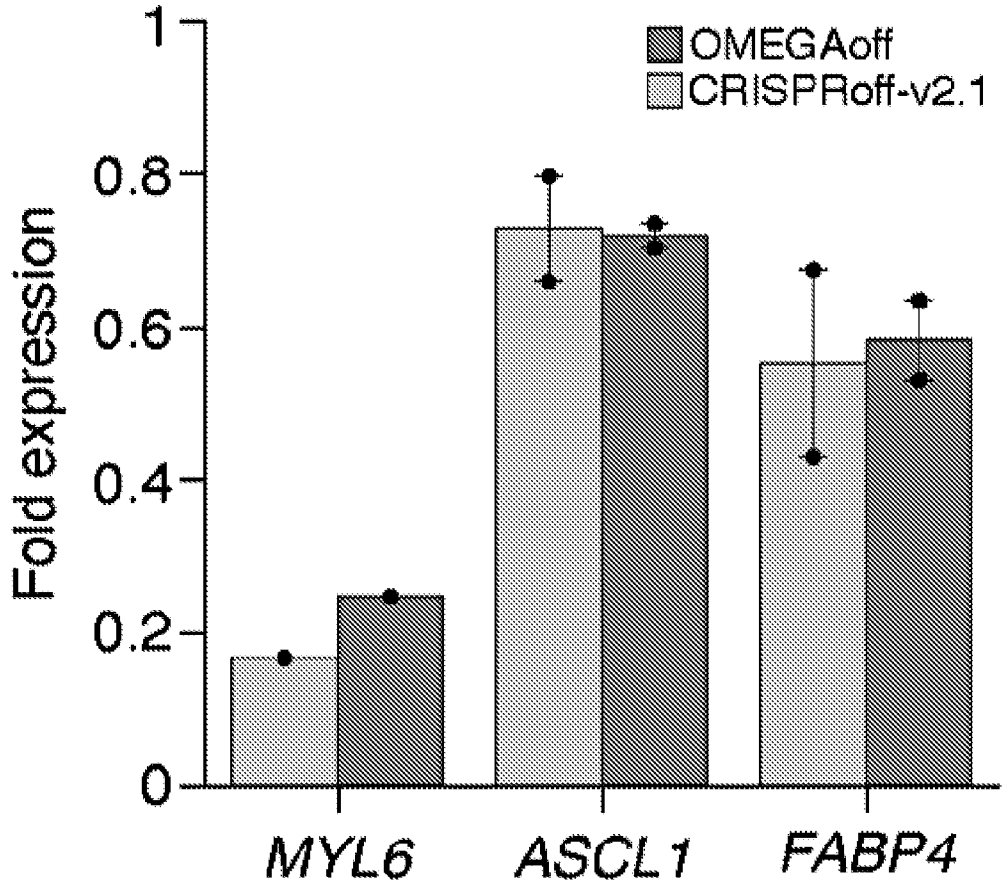

OMEGAoff
CRISPRoff-v2.1
Fold expression
1
0.8
0.6
0.4
0.2
0
MYL6
ASCL1
FABP4

FIG. 120C