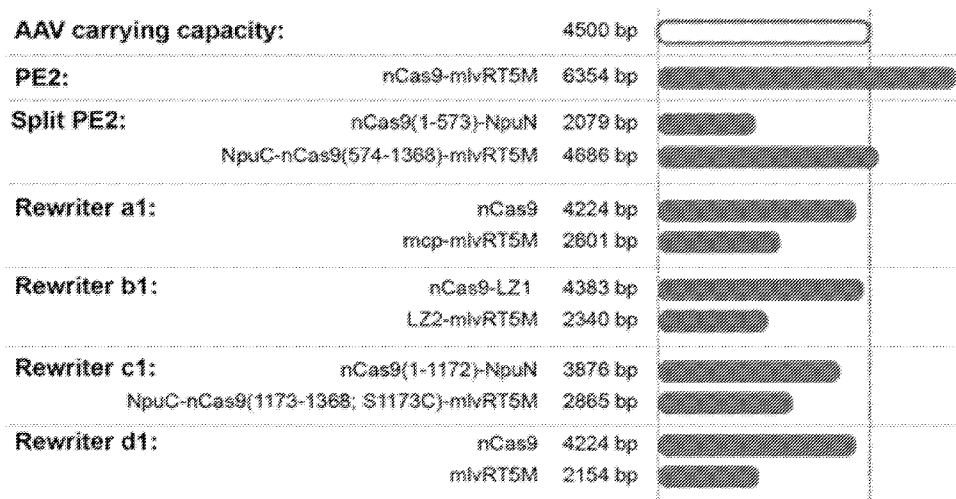




- (51) International Patent Classification:
CI2N 9/12 (2006.01) *CI2N 9/22* (2006.01)
- (21) International Application Number:
 PCT/US2021/023041
- (22) International Filing Date:
 18 March 2021 (18.03.2021)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
 62/992,032 19 March 2020 (19.03.2020) US
 63/055,829 23 July 2020 (23.07.2020) US
 63/153,161 24 February 2021 (24.02.2021) US
- (71) Applicant: **REWRITE THERAPEUTICS, INC.**
 [US/US]; 1250 Powell Street, Apartment 4, Emeryville,
 California 94608 (US).
- (72) Inventor: **HALPERIN, Schaked Omer**; 1250 Powell
 Street, Apartment 4, Emeryville, California 94608 (US).
- (74) Agent: **HECKER, Peter**; WILSON SONSINI
 GOODRICH & ROSATI, 650 Page Mill Road, Palo Alto,
 California 94304-1050 (US).
- (81) Designated States (*unless otherwise indicated, for every
 kind of national protection available*): AE, AG, AL, AM,
 AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
 CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
 DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
 HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN,
 KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
 ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO,
 NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
 SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
 TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: METHODS AND COMPOSITIONS FOR DIRECTED GENOME EDITING

FIG. 19A



(57) Abstract: Provided herein are compositions and methods for increasing editing efficiency of a target nucleic acid. A composition may comprise a guide nucleic acid, a Cas9 nickase, or a reverse transcriptase. The reverse transcriptase may be fused to the Cas9 nickase. The reverse transcriptase may heterodimerize with the Cas9 nickase. The reverse transcriptase may bind to a guide nucleic acid. The reverse transcriptase may be engineered to increase processivity. The guide nucleic acid may be engineered to facilitate synthesis or editing of a sequence. The guide nucleic acid, Cas9 nickase, and reverse transcriptase may be engineered to fit within AAV vectors. The guide nucleic acid may comprise a region that binds to another region on the guide nucleic acid to improve gene editing.



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

Published:

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

METHODS AND COMPOSITIONS FOR DIRECTED GENOME EDITING**CROSS-REFERENCE**

[0001] This application claims the benefit of U.S. Provisional Application No. 62/992,032 filed March 19, 2020, U.S. Provisional Application No. 63/055,829 filed July 23, 2020, and U.S. Provisional Application No. 63/153,161 filed February 24, 2021, all of which are incorporated herein by reference in their entireties.

INCORPORATION BY REFERENCE OF SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on March 16, 2021, is named 56385-701_601_SL.txt and is 609,554 bytes in size.

BACKGROUND

[0003] Cas-directed genome editing techniques that introduce double-strand breaks in a target nucleic acid frequently result in undesired products including sequence translocations, insertions, deletions, and activation of DNA damage repair, cell cycle arrest, or apoptosis functions of p53. Editing techniques that introduce single-strand breaks may be limited in the type and size of permissible mutations or may have limited editing efficiency. There is a need for genome editing techniques with improved accuracy, efficiency, and versatility.

SUMMARY

[0004] In various aspects, the present disclosure provides a method of increasing gene editing efficiency in a cell having a low deoxynucleoside triphosphate (dNTP) concentration and comprising a DNA polymerase, the method comprising: increasing the dNTP concentration in the cell, relative to a baseline dNTP concentration. In various aspects, increasing the dNTP concentration in the cell comprises inhibiting a deoxynucleotide triphosphate triphosphohydrolase in the cell. In various aspects, the deoxynucleotide triphosphate triphosphohydrolase comprises SAM domain and HD domain-containing protein 1 (SAMHD1). In various aspects, inhibiting SAMHD1 comprises contacting the SAMHD1 with a Vpx protein, or expressing the Vpx protein in the cell. In various aspects, inhibiting SAMHD1 comprises contacting the SAMHD1 with a BGLF4 protein, or expressing the BGLF4 protein in the cell. In various aspects, inhibiting SAMHD1 comprises contacting an mRNA encoding the SAMHD1 with a microRNA or siRNA that hybridizes to the mRNA, or expressing the microRNA or siRNA in the cell. In various aspects, inhibiting SAMHD1 comprises contacting the SAMHD1

with a small molecule SAMHD1 inhibitor. In various aspects, increasing the dNTP concentration in the cell comprises administering nucleosides or nucleotides (e.g., dNs, dNMPs, or NTPs) to the cell. In various aspects, administering dNTPs to the cell comprises administering the nucleosides or nucleotides to a subject comprising the cell. In various aspects, the administration is oral or by injection. In various aspects, increasing the dNTP concentration in the cell comprises delivering a dNTP synthetic enzyme to the cell. In various aspects, the dNTP synthetic enzyme comprises a kinase. In various aspects, the kinase comprises a nucleoside kinase, deoxynucleoside kinase, deoxynucleoside monophosphate kinase, or deoxynucleotide diphosphate kinase. In various aspects, the DNA polymerase comprises a reverse transcriptase. In various aspects, the cell further comprises a Cas9 programmable nuclease, a guide nucleic acid, or a combination thereof. In various aspects, the low dNTP concentration comprises a dNTP concentration found in a nondividing cell. In various aspects, the low dNTP concentration is less than a dNTP concentration found in an activated peripheral blood mononuclear cell. In various aspects, the low dNTP concentration comprises a dNTP concentration below 1 micromolar. In various aspects, the increasing the dNTP concentration comprises increasing the dNTP concentration by at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, or more, relative to the baseline dNTP measurement. In various aspects, the dNTP concentration comprises a deoxyadenosine triphosphate (dATP) concentration, a deoxycytidine triphosphate (dCTP) concentration, a deoxyguanosine triphosphate (dGTP) concentration, or a deoxythymidine triphosphate (dTTP) concentration, or any combination thereof.

[0005] In various aspects, the present disclosure provides a method of increasing genome editing efficiency comprising expressing a Vpx protein in a cell expressing the composition as described herein or the guide nucleic acid as described herein.

[0006] In various aspects, the present disclosure provides a method of increasing genome editing efficiency by increasing the dNTP concentration in a cell. In various aspects, the present disclosure provides a method of increasing genome editing efficiency comprising inhibiting SAMHD1 in a cell expressing a Cas9 programmable nuclease, a reverse transcriptase, and a guide nucleic acid. In various aspects, the present disclosure provides a method of increasing genome editing efficiency comprising increasing a dNTP concentration (e.g. by inhibiting SAMHD1) in a cell expressing a Cas9 programmable nuclease, a reverse transcriptase, and a guide nucleic acid.

[0007] In some aspects, inhibiting SAMHD1 comprises expressing a Vpx protein in the cell. In some aspects, inhibiting SAMHD1 comprises expressing a microRNA against SAMHD1 in the

cell. In some aspects, inhibiting SAMHD1 comprises treating the cell with a small molecule SAMHD1 inhibitor.

[0008] In various aspects, the present disclosure provides a composition comprising a Cas nickase and a reverse transcriptase, wherein the Cas nickase and the reverse transcriptase are separate polypeptide chains and the Cas nickase and the reverse transcriptase form a Cas-reverse transcriptase heterodimer.

[0009] In some aspects, the Cas-reverse transcriptase heterodimer comprises a first heterodimer domain fused to the Cas nickase and a second heterodimer domain fused to the reverse transcriptase, wherein the first heterodimer domain binds the second heterodimer domain to form a heterodimer. In some aspects, the first heterodimer domain is a leucine zipper and the second heterodimer domain is a leucine zipper. In some aspects, the reverse transcriptase comprises a sequence having at least 80% sequence identity to of any one of SEQ ID NO: 3 – SEQ ID NO: 22 or a fragment thereof. In some aspects, the reverse transcriptase comprises a domain from a non-long terminal repeat retrotransposable element fused to a Cas nickase. In some aspects, the reverse transcriptase comprises a sequence from a bacterial group II intron fused to a Cas nickase. In some aspects, the reverse transcriptase comprises a domain from a retroviral gag-pol polyprotein fused to a Cas nickase.

[0010] Disclosed herein are compositions comprising a Cas nickase and a reverse transcriptase, wherein the Cas nickase and the reverse transcriptase comprise separate polypeptide chains, and wherein the Cas nickase and reverse transcriptase are not engineered to heterodimerize.

[0011] In various aspects, the present disclosure provides a composition comprising a Cas nickase, a reverse transcriptase, and a guide nucleic acid, wherein a first polypeptide comprises the Cas nickase and a second polypeptide comprises the reverse transcriptase and the guide nucleic acid binds to the Cas nickase and the reverse transcriptase.

[0012] Some aspects comprise a guide nucleic acid that forms a complex with the Cas nickase, wherein, upon complex formation, the Cas nickase is capable of introducing a single-strand break at a target site in a target nucleic acid. In some aspects, the target nucleic acid comprises a CFTR nucleic acid, a USH2A nucleic acid, an ABCA4 nucleic acid, an ATP7B nucleic acid, or an HTT nucleic acid.

[0013] In some aspects, the reverse transcriptase comprises an mcp peptide. In some aspects, the reverse transcriptase comprises a loop region. In some aspects, the loop region is a 2a loop or a 3a loop. In some aspects, the loop region is a 2a loop. In some aspects, the loop region is a 3a loop. In some aspects, the guide nucleic acid comprises a MS2 hairpin.

[0014] In various aspects, the present disclosure provides a composition comprising a reverse transcriptase with a sequence having at least 80% sequence identity to of any one of SEQ ID NO: 3 – SEQ ID NO: 22 or a fragment thereof fused to a Cas nickase.

[0015] In various aspects, the present disclosure provides a composition comprising a reverse transcriptase comprising a domain from a non-long terminal repeat retrotransposable element fused to a Cas nickase.

[0016] In various aspects, the present disclosure provides a composition comprising a reverse transcriptase comprising a sequence from a bacterial group II intron fused to a Cas nickase.

[0017] In various aspects, the present disclosure provides a composition comprising a reverse transcriptase comprising a domain from a retroviral gag-pol polyprotein fused to a Cas nickase.

[0018] In some aspects, the composition comprises a guide nucleic acid that complexes with the Cas nickase and the reverse transcriptase, wherein, upon complex formation, the Cas nickase is capable of introducing a single-strand break at a target site in a target nucleic acid. In some aspects, the composition comprises a nuclear localization signal fused to the Cas nickase or the reverse transcriptase. In some aspects, the reverse transcriptase is a truncated reverse transcriptase. In some aspects, the reverse transcriptase has an increased processivity as compared to a native reverse transcriptase. In some aspects, the reverse transcriptase has increased processivity compared to mlvRT. In some aspects, the reverse transcriptase edits a longer window length in a target sequence compared to mlvRT. In some aspects, the reverse transcriptase has decreased immunogenicity compared to mlvRT. In some aspects, the reverse transcriptase has improved delivery to a cell compared to mlvRT. In some aspects, the reverse transcriptase polymerizes 20 or more, 40 or more, 45 or more, 50 or more, 60 or more, 81 or more, 100 or more, 500 or more, or 1000 or more nucleotides in a single binding event.

[0019] In various aspects, the present disclosure provides a composition comprising a Cas nickase and a reverse transcriptase, or at least one polynucleotide encoding the Cas nickase and reverse transcriptase, wherein at least part of the Cas nickase and the reverse transcriptase are included in at least 2 separate polypeptide chains. In some aspects, the at least 2 separate polypeptide chains comprise separate polypeptide chains comprising heterodimer domains that bind one another. In some aspects, the at least 2 separate polypeptide chains comprise separate polypeptide chains comprising inteins that bind one another, and the Cas nickase comprises a mutation at amino acid position 1030 or after amino acid position 1030, the mutation comprising a point mutation to a cysteine, threonine, alanine, or serine, or an insertion of a cysteine, threonine, alanine, or serine. In some aspects, the at least 2 separate polypeptide chains comprise the separate polypeptide chains comprising heterodimer domains that bind one another. In some

aspects, the separate polypeptide chains comprise fusion proteins comprising the heterodimer domains. In some aspects, the heterodimer domains comprise leucine zippers, PDZ domains, streptavidin and streptavidin binding protein, foldon domains, hydrophobic polypeptides, an antibody that binds the Cas nickase, or an antibody that binds the reverse transcriptase, or one or more binding fragments thereof. In some aspects, the heterodimer domains comprise a first heterodimer domain and a second heterodimer domain, the Cas nickase comprising the first heterodimer domain and the reverse transcriptase comprising the second heterodimer domain. In some aspects, the first heterodimer domain is fused to an amino or carboxy end of the Cas nickase, and the second heterodimer domain is fused to an amino or carboxy end of the reverse transcriptase. In some aspects, the first heterodimer domain comprises a first leucine zipper, and wherein the second heterodimer domain comprises a second leucine zipper. In some aspects, the at least 2 separate polypeptide chains comprise the separate polypeptide chains comprising the inteins that bind one another, and the Cas nickase comprises the mutation at amino acid position 1030 or after amino acid position 1030, the mutation comprising a point mutation to a cysteine, threonine, alanine, or serine, or an insertion of a cysteine, threonine, alanine, or serine. In some aspects, the point mutation is to a cysteine, or the insertion is of a cysteine. In some aspects, the point mutation is to a threonine, or the insertion is of a threonine. In some aspects, the point mutation is to an alanine, or the insertion is of an alanine. In some aspects, the point mutation is to a serine, or the insertion is of a serine. In some aspects, the mutation comprises the point mutation, wherein the point mutation is at amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, R1212, or a range defined by any two of the aforementioned amino acid positions. In some aspects, the mutation comprises the point mutation, wherein the point mutation is at amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212. In some aspects, the mutation comprises the insertion mutation, wherein the insertion mutation is immediately upstream or downstream of amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, R1212, or a range defined by any two of the aforementioned amino acid positions. In some aspects, the mutation comprises the insertion mutation, wherein the insertion mutation is immediately upstream or downstream of amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212. In some aspects, the inteins comprise a first intein and a second intein, and wherein the Cas nickase comprises a first segment comprising the first intein, and a second segment comprising the mutation and the second intein. Some aspects include a guide nucleic acid that binds to the Cas nickase or the reverse transcriptase. In some aspects, the Cas nickase of the complex introduces a single-strand

break at a target site in a target nucleic acid. In some aspects, the Cas nickase comprises a Cas9 nickase or a variant thereof. In some aspects, the Cas9 nickase or variant thereof comprises an *S. Pyogenes* Cas9 nickase or a variant thereof. In some aspects, the reverse transcriptase comprises a Moloney leukemia virus reverse transcriptase (mlvRT) or a variant thereof. In some aspects, the reverse transcriptase comprises a point mutation at position P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524. In some aspects, the reverse transcriptase comprises a point mutation comprising P51L, S67R, Q84A, L139P, Q221R, V223A, V223M, T197A, D653N, T664N, L671P, L435G, H204R, or D524A. In some aspects, the reverse transcriptase comprises a point mutation at amino acid position Q84, L139, Q221, V223, T664, or L671. In some aspects, the reverse transcriptase comprises a point mutation comprising S67R, Q84A, L139P, Q221R, V223A, V223M, T664N, L671P, or D524A. In some aspects, the composition comprises the Cas nickase and the reverse transcriptase, and wherein the at least 2 separate polypeptide chains are 2 separate polypeptide chains. In some aspects, the composition comprises the Cas nickase and the reverse transcriptase, and wherein the at least 2 separate polypeptide chains comprise a first polypeptide chain comprising a first part of the Cas nickase, and a second polypeptide chain comprising a second part of the Cas nickase and the reverse transcriptase. Some aspects include the at least one polynucleotide encoding the Cas nickase and reverse transcriptase. In some aspects, the at least one polynucleotide encoding the Cas nickase and reverse transcriptase comprises a first polynucleotide encoding a first part of the Cas nickase, and a second polynucleotide encoding a second part of the Cas nickase and the reverse transcriptase. Some aspects include at least one adeno-associated virus comprising the at least one polynucleotide. In some aspects, the composition is produced by a cell.

[0020] In various aspects, the present disclosure provides a composition comprising a Cas nickase and a reverse transcriptase, or at least one polynucleotide encoding the Cas nickase and reverse transcriptase, wherein at least part of the Cas nickase and the reverse transcriptase are included in separate polypeptide chains. In some aspects, the Cas nickase or the reverse transcriptase comprise a first leucine zipper. In some aspects, the Cas9 nickase comprises an *S. Pyogenes* Cas9 nickase, or a variant thereof, and a point mutation at amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212, or an insertion mutation immediately upstream or downstream of amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212. In some aspects, the reverse transcriptase comprises a Moloney leukemia virus reverse transcriptase (mlvRT), or a variant thereof, and a point mutation at amino acid position Q84, L139, Q221, V223, T664, or L671. In some aspects, the separate polypeptide chains comprise heterodimer domains. In some aspects,

the Cas nickase and the reverse transcriptase form a heterodimer comprising a first heterodimer domain fused to the Cas nickase and a second heterodimer domain fused to the reverse transcriptase, wherein the first heterodimer domain binds to the second heterodimer domain to form the heterodimer. In some aspects, the first heterodimer domain comprises the first leucine zipper. In some aspects, the second heterodimer domain comprises a second leucine zipper. In some aspects, the reverse transcriptase comprises a sequence having at least 80% sequence identity to of any one of SEQ ID NO: 3 – SEQ ID NO: 22 or SEQ ID NO: 40 – SEQ ID NO: 80, or a fragment thereof. In some aspects, the reverse transcriptase comprises a domain from a non-long terminal repeat retrotransposable element fused to part of the Cas nickase, a sequence from a bacterial group II intron fused to part of the Cas nickase, or a domain from a retroviral gag-pol polyprotein fused to part of the Cas nickase. Some aspects include a guide nucleic acid that binds to the Cas nickase or the reverse transcriptase. In some aspects, the Cas nickase of the complex introduces a single-strand break at a target site in a target nucleic acid. In some aspects, the guide nucleic acid comprises a hairpin. In some aspects, the hairpin comprises an MS2 hairpin. In some aspects, the reverse transcriptase comprises a modified reverse transcriptase comprising a hairpin binding domain. In some aspects, the reverse transcriptase comprises a modified reverse transcriptase comprising an MS2 coat protein (MCP) peptide. In some aspects, the reverse transcriptase comprises a loop region. In some aspects, the Cas9 nickase comprises a point mutation or an insertion mutation in a C-terminal half of the Cas9 nickase. In some aspects, the point mutation in the C-terminal half of the Cas9 nickase is to a cysteine, serine, threonine, or alanine; or wherein the insertion mutation is a cysteine insertion serine insertion, threonine insertion, or alanine insertion. In some aspects, the Cas9 nickase comprises the point mutation in the C-terminal half of the Cas9 nickase. In some aspects, the Cas9 nickase comprises the insertion mutation in the C-terminal half of the Cas9 nickase. In some aspects, the Cas9 nickase comprises a first segment comprising a first intein, and a second segment comprising the point mutation or insertion mutation and a second intein. In some aspects, the Cas9 nickase comprises the *S. Pyogenes* Cas9 nickase or variant thereof. In some aspects, the Cas9 nickase comprises the point mutation at amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212, or the insertion mutation immediately upstream or downstream of amino acid position D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212. In some aspects, the Cas nickase or the reverse transcriptase comprises a nuclear localization signal. In some aspects, the reverse transcriptase is a truncated reverse transcriptase. In some aspects, the reverse transcriptase comprises the mlvRT or variant thereof. In some aspects, the reverse transcriptase comprises the point mutation at amino acid

position Q84, L139, Q221, V223, T664, or L671. In some aspects, the reverse transcriptase comprises a point mutation comprising P51L, S67R, Q84A, L139P, Q221R, V223A, V223M, T197A, D653N, T664N, L671P, L435G, H204R, or D524A. In some aspects, the reverse transcriptase comprises a point mutation comprising S67R, Q84A, L139P, Q221R, V223A, V223M, T664N, L671P, or D524A. In some aspects, the Cas nickase and the reverse transcriptase comprise separate polypeptide chains. In some aspects, the composition is produced by a cell. Some aspects include the at least one polynucleotide encoding the Cas nickase and reverse transcriptase. Some aspects include at least one adeno-associated virus comprising the at least one polynucleotide.

[0021] In various aspects, the present disclosure provides a guide nucleic acid comprising: a spacer reverse complementary to a first region of a target nucleic acid, a scaffold configured to bind to a Cas nickase, a reverse transcriptase template encoding a sequence to be inserted into the target nucleic acid, and a first strand primer binding site reverse complementary to a second region of the target nucleic acid.

[0022] In some aspects, the guide nucleic acid further comprises a second strand primer comprising a sequence of a region of the reverse transcriptase template. In some aspects, the first region of the target nucleic acid is on a first strand of the target nucleic acid and the second region of the target nucleic acid is on a second strand of the target nucleic acid. In some aspects, all or part of the first region of the target nucleic acid is reverse complementary to all or part of the second region of the target nucleic acid. In some aspects, the guide nucleic acid further comprises a cleavable sequence at the 3' end of the guide nucleic acid. In some aspects, the cleavable sequence is a ribozyme cleavable sequence. In some aspects, the cleavable sequence is a tRNA cleavable sequence. In some aspects, the first strand primer binding site is configured to hybridize to the second region of the target nucleic acid, and wherein the reverse transcriptase template is configured to serve as a template for reverse transcription from a 3' end of the second region of the target nucleic acid. In some aspects, the second strand primer is configured to serve as a primer for transcription from a template reverse complementary to the reverse transcriptase template. In some aspects, a first synthesized strand serves as a template for synthesis of a second strand from the second strand primer. In some aspects, a Velcro region that hybridizes to a region of the reverse transcriptase template region.

[0023] In various aspects, the present disclosure provides a composition comprising a first guide nucleic acid comprising the guide as described herein and a second guide nucleic acid.

[0024] In some aspects, the second guide nucleic acid comprises the guide as described herein. In some aspects, the first guide nucleic acid binds to a first Cas nickase, and the second guide

nucleic acid binds to a second Cas nickase. In some aspects, a first spacer of the first guide nucleic acid binds a first Cas nickase, a second spacer of the second guide nucleic acid binds a second Cas nickase, a first scaffold of the first guide nucleic acid binds the second Cas nickase, and a second scaffold of the second guide nucleic acid binds the first Cas nickase. In some aspects, the first guide nucleic acid comprises a first linker and the second guide nucleic acid comprises a second linker, wherein the first linker hybridizes to the second linker.

[0025] In various aspects, the present disclosure provides a guide nucleic acid comprising: a spacer reverse complementary to a first region of a target nucleic acid; a scaffold configured to bind to a Cas nuclease; a reverse transcriptase template encoding a sequence to be inserted into the target nucleic acid; a first strand primer binding site reverse complementary to a second region of the target nucleic acid; and at least one of: i. a gRNA positioning system (GPS) region and a GPS binding site that hybridizes to the GPS region, ii. a modification in the reverse transcriptase template that disrupts a protospacer adjacent motif (PAM) sequence in the target nucleic acid, iii. a modification in the reverse transcriptase template that disrupts a track of at least 4 consecutive nucleotides of the same base in the target nucleic acid, or iv. a second strand primer comprising a sequence of a region of the reverse transcriptase template. In various aspects, the present disclosure provides a guide nucleic acid comprising: a spacer reverse complementary to a first region of a target nucleic acid; a scaffold configured to bind to a Cas nuclease; a reverse transcriptase template encoding a sequence to be inserted into the target nucleic acid; a first strand primer binding site reverse complementary to a second region of the target nucleic acid; and at least one of: i. a gRNA positioning system (GPS) region and a GPS binding site that hybridizes to the GPS region, ii. a modification in the reverse transcriptase template that disrupts a track of at least 4 consecutive nucleotides of the same base in the target nucleic acid, or iii. a second strand primer comprising a sequence of a region of the reverse transcriptase template. Some aspects include the second strand primer. In some aspects, the second strand primer is configured to serve as a primer for transcription from a template reverse complementary to the reverse transcriptase template. In some aspects, a first synthesized strand serves as a template for synthesis of a second strand from the second strand primer. In some aspects, the first region of the target nucleic acid is on a first strand of the target nucleic acid and the second region of the target nucleic acid is on a second strand of the target nucleic acid. In some aspects, all or part of the first region of the target nucleic acid is reverse complementary to all or part of the second region of the target nucleic acid. Some aspects include a ribozyme cleavable sequence at a 3' end of the guide nucleic acid. Some aspects include a tRNA cleavable sequence at a 3' end of the guide nucleic acid. In some aspects, the first strand primer binding

site is configured to hybridize to the second region of the target nucleic acid, and wherein the reverse transcriptase template is configured to serve as a template for reverse transcription from a 3' end of the second region of the target nucleic acid. Some aspects include the GPS region and the GPS binding site. In some aspects, the GPS region and the GPS binding site together comprise a region of the guide nucleic acid that binds to another region on the guide nucleic acid to affect a conformational change in the guide nucleic acid and improve gene editing. In some aspects, the hybridization of the GPS region and the GPS binding site conformationally changes the guide nucleic acid, and improves editing efficiency as compared to a guide nucleic acid without the GPS region or GPS binding site. In some aspects, the reverse transcriptase template region comprises the GPS binding site. In some aspects, the GPS binding site is 5' of the first strand primer binding site. In some aspects, the GPS binding site is 3' of the first strand primer binding site. In some aspects, the GPS region is 5' of the reverse transcriptase template. In some aspects, the GPS region is 3' of the reverse transcriptase template. In some aspects, the GPS region is 5' of the scaffold. In some aspects, the GPS region is 5-100 nucleotides in length. In some aspects, the GPS binding site is at least 50% complementary to the GPS region. In some aspects, the target nucleic acid comprises a CFTR nucleic acid, a USH2A nucleic acid, an ABCA4 nucleic acid, an ATP7B nucleic acid, or an HTT nucleic acid. In some aspects, the spacer comprises a nucleic acid sequence at least 85% identical to any one of SEQ ID NOs: 96-119. Some aspects include the modification in the reverse transcriptase template that disrupts the PAM sequence in the target nucleic acid. In some aspects, the PAM sequence comprises a 2-6 base pair nucleic acid sequence recognized by the Cas nuclease. In some aspects, the modification in the reverse transcriptase template that disrupts the PAM sequence in the target nucleic acid improves gene editing relative to a guide nucleic acid without the modification. Some aspects include the modification in the reverse transcriptase template that disrupts the track of at least 4 consecutive nucleotides of the same base in the target nucleic acid. In some aspects, the track of at least 4 consecutive nucleotides of the same base comprise a polyA track. In some aspects, the modification in the reverse transcriptase template that disrupts the track of at least 4 consecutive nucleotides of the same base in the target nucleic acid improves gene editing relative to a guide nucleic acid without the modification. In some aspects, the Cas nuclease comprises a Cas nickase. In some aspects, the guide nucleic acid comprises a guide RNA. Some aspects include a gene editing method comprising delivering a composition comprising the guide nucleic acid to a cell. In some aspects, the composition comprises a viral vector comprising the guide nucleic acid. Some aspects include the GPS region that hybridizes to the GPS binding site on the second guide nucleic acid.

[0026] In various aspects, the present disclosure provides a method of increasing genome editing efficiency comprising delivering an Orf1p to a cell expressing the composition as described herein or the guide nucleic acid as described herein.

[0027] In various aspects, the present disclosure provides a nucleic acid comprising nucleotide sequence encoding the composition as described herein or the guide nucleic acid as described herein.

[0028] In various aspects, the present disclosure provides a viral vector comprising the nucleic acid as described herein.

[0029] In various aspects, the present disclosure provides a cell comprising the composition as described herein, the guide nucleic acid as described herein, the nucleic acid as described herein, or the viral vector as described herein.

[0030] In some aspects, the cell is a prokaryotic cell. In some aspects, the cell is a eukaryotic cell.

[0031] In some aspects, the present disclosure provides a composition comprising a Cas9 programmable nuclease comprising one or more point mutations or insertion mutations that enable or improve intein catalysis. In various aspects, the present disclosure provides a composition comprising a Cas9 programmable nuclease, wherein the Cas9 programmable nuclease comprises a cysteine point mutation located in a C-terminal half of the Cas9 programmable nuclease. In various aspects, the present disclosure provides a composition comprising a Cas9 programmable nuclease, wherein the Cas9 programmable nuclease comprises an insertion mutation (e.g. a cysteine insertion mutation) located in a C-terminal half of the Cas9 programmable nuclease. The point mutation may be a cysteine point mutation, a serine point mutation, a threonine point mutation, or an alanine point mutation. The insertion mutation may be a cysteine insertion mutation, a serine insertion mutation, a threonine insertion mutation, or an alanine insertion mutation.

[0032] In some aspects, the Cas9 programmable nuclease is a Cas9 nickase. In some aspects, the Cas9 programmable nuclease is an *S. Pyogenes* Cas9. In some aspects, the point mutation is located at D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212 of the *S. Pyogenes* Cas9. In some aspects, the insertion mutation is immediately upstream of D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212 of the *S. Pyogenes* Cas9. In some aspects, the cysteine point mutation is located at S1173C, D1079C, or D1180C of the *S. Pyogenes* Cas9. In some aspects, the cysteine insertion mutation is located at 1173C, 1079C, or 1180C of the *S. Pyogenes* Cas9. In some aspects, the Cas9 programmable

nuclease comprises a sequence of any one of SEQ ID NO: 85 – SEQ ID NO: 87 or SEQ ID NO: 90 – SEQ ID NO: 92.

[0033] In some aspects, the Cas9 programmable nuclease is expressed as two or more segments. In some aspects, a first segment of the two or more segments comprise an N-terminal portion of the Cas9 programmable nuclease and a first intein, and wherein a second segment of the two or more segments comprise a C-terminal portion of the Cas9 programmable nuclease and a second intein. In some aspects, the cysteine point mutation is located at the N-terminus of the C-terminal portion of the Cas9 programmable nuclease. In some aspects, the cysteine insertion mutation is located at the N-terminus of the C-terminal portion of the Cas9 programmable nuclease. In some aspects, the first intein is fused to the C-terminus of the N-terminal portion of the Cas9 programmable nuclease, and wherein the second intein is fused to the N-terminus of the C-terminal portion of the Cas9 programmable nuclease. In some aspects, the first segment comprises a sequence of SEQ ID NO: 90, and wherein the second segment comprises a sequence of SEQ ID NO: 91. The second segment of the two or more segments may comprise a reverse transcriptase fused to the C-terminal portion of the Cas9 programmable nuclease. The reverse transcriptase may comprise an N-terminus fused to a C-terminus of the C-terminal portion of the Cas9 programmable nuclease. The reverse transcriptase may comprise an mlvRT, or a variant thereof.

[0034] Disclosed herein are methods of optimizing genome editing efficiency comprising performing genome editing with a Moloney leukemia virus reverse transcriptase (mlvRT) that is modified to increase its catalytic efficiency in low dNTP concentrations, (e.g. modified to decrease its K_m for dNTPs). Disclosed herein are methods of optimizing genome editing efficiency in a limiting dNTP condition, comprising performing genome editing with a Moloney leukemia virus reverse transcriptase (mlvRT), or a variant thereof, comprising a point mutation at position 221 or 223 of the reverse transcriptase. The mlvRT or variant thereof may comprise a point mutation at position 221. The point mutation at position 221 may comprise Q221R. The mlvRT or variant thereof may comprise a point mutation at position 223. The point mutation at position 223 may comprise V223A. The point mutation at position 223 may comprise V223M.

[0035] The Cas nickase and RT may be encoded by polynucleotides. Disclosed herein are AAVs comprising the polynucleotides. At least part of the Cas nickase and RT may be encompassed or comprised within separate AAVs. Disclosed herein are AAVs comprising a first AAV comprising a first polynucleotide encoding a Cas or Cas9 component, and a second AAV comprising a second polynucleotide encoding a RT component. The AAVs may comprise AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11,

AAV12, AAV-DJ, AAV-DJ/8, AAV-Rh10, AAV-Rh74, AAV-retro, AAV-PHP.B, AAV8-PHP.eB, or AAV-PHP.S, or a combination of thereof.

[0036] Disclosed herein are methods for screening or identifying an improved reverse transcriptase (RT), comprising: overexpressing SAMHD1, or expressing a mutant SAMHD1 that has been mutated to prevent phosphorylation of a residue of the mutant SAMHD1, in cells; identifying an RT activity in the cells; and based on the RT activity, identifying the RT as an improved RT.

[0037] Disclosed herein are systems comprising an RNA or polynucleotide comprising a spacer, a reverse transcriptase template comprising a desired edit, and a primer binding site, in which the primer binding site binds to a nucleic acid that does not comprise any part of the region of the nucleic acid targeted or bound by the spacer or the nucleic acid reverse complementary to the nucleic acid targeted or bound by the spacer.

[0038] Disclosed herein are systems comprising: a first guide nucleic acid comprising: a spacer reverse complementary to a first region of a target nucleic acid; a scaffold configured to bind to a Cas nuclease; a reverse transcriptase template encoding a sequence to be inserted into the target nucleic acid; a first strand primer binding site that binds to a region of the target nucleic acid that does not comprise any part of the first region, and that does not comprise any part of a reverse complement of the first region; and a GPS region that hybridizes to a GPS binding site on a second guide nucleic acid. Disclosed herein are systems comprising: a first guide nucleic acid comprising: a spacer reverse complementary to a first region of a target nucleic acid; a scaffold configured to bind to a Cas nuclease; a reverse transcriptase template encoding a sequence to be inserted into the target nucleic acid; and a first strand primer binding site that binds to a region of the target nucleic acid that does not comprise any part of the first region, and that does not comprise any part of a reverse complement of the first region. Some aspects include a GPS region that hybridizes to a GPS binding site on a second guide nucleic acid. Some aspects include the second guide nucleic acid. The second guide nucleic acid may include the GPS binding site. In some aspects, the second guide nucleic acid comprises a second spacer reverse complementary to another region of the target nucleic acid. The second guide nucleic acid may bring the primer binding site into proximity or contact with a genomic flap.

INCORPORATION BY REFERENCE

[0039] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0040] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0041] FIG. 1 shows the editing efficiency of a fused Cas9 nickase (nCas9) reverse transcriptase (RT) construct (“nCas9-mlvRT”) comprising an nCas9 and a Moloney leukemia virus RT (“mlvRT”), and a split nCas9-LZ1 and LZ2-mlvRT construct (“mlvRT Split Stitch”). The split nCas9-LZ1 and LZ2-mlvRT construct comprises a nCas9-LZ1 (SEQ ID NO: 1, NLS-SpCas9(H840A)-NLS-EE12RR345L(leucine zipper)) and a LZ2-mlvRT (SEQ ID NO: 2, RR12EE345L(leucine zipper)-mlvRTv(nCas9-mlvRT(D200N, L603W, T306K, W313F, T330P)-NLS) on discrete polypeptide chains. The nCas9-LZ1 comprises a SpCas9 (SEQ ID NO: 32) and a C-terminal leucine zipper (SEQ ID NO: 23) that heterodimerizes with the LZ2-mlvRT comprising a mlvRT (SEQ ID NO: 13) and an N-terminal leucine zipper (SEQ ID NO: 24) through the leucine zippers. Schematics of the nCas9-mlvRT constructs are provided at top.

[0042] FIG. 2 shows the editing efficiency of a fused nCas9-RT construct (“nCas9-mlvRT”) and a split nCas9 and mcp-mlvRT construct (“mcp-mlvRTv”) comprising an nCas9 and a mcp peptide fused to reverse transcriptase (SEQ ID NO: 27). The mcp peptide interacts with MS2 RNA hairpins. Efficiency of the split nCas9 and mcp-mlvRT construct was tested with different guide RNA (gRNA) constructs including gRNA 2.0 (SEQ ID NO: 31), a gRNA with a long MS2 hairpin (SEQ ID NO: 28), “gRNA-1xlongMS2”), a gRNA with a short MS2 hairpin (SEQ ID NO: 29, “gRNA-1xshortMS2”), or a gRNA with two short MS2 hairpins (SEQ ID NO: 30, “gRNA-2xshortMS2”).

[0043] FIG. 3 shows the editing efficiency of different split nCas9-RT constructs comprising modified reverse transcriptases with increased transcriptional processivity. Constructs comprising nCas9 and reverse transcriptases from either *geobacillus stereothermophilus* (Gsl-IICRT, SEQ ID NO: 3), *Eubacterium Rectale* (ErRT, SEQ ID NO: 4), and amino acids 116-1016 from the R2 polyprotein (R2(116-1016), SEQ ID NO: 7) were tested. A schematic of the Gsl-IICRT reverse transcriptase (“StitchRT”) is shown compared to the mlvRT reverse transcriptase used in **FIG. 1** and **FIG. 2**.

[0044] FIG. 4A illustrates a method for genome editing using an engineered gRNA of the present disclosure (“Stitch Guide”). A nCas9-RT construct complexed with a gRNA is recruited to a target site of a target nucleic acid by hybridization of a spacer of the gRNA to the target site.

The nCas9 nicks a strand of a target nucleic acid at a target site. A first strand primer binding site of the gRNA hybridizes to a flap 5' of the nick. The RT polymerizes from the 3' end of the flap using a reverse transcriptase template region of the gRNA as a template. A second strand primer ("2nd strand primer") at the 3' end of the gRNA hybridizes to the 3' end of the newly synthesized DNA strand. The 4-200 bp second strand primer region acts as an RNA primer for synthesis of a second DNA strand. The RT polymerizes from the 3' end of the gRNA using the newly synthesized DNA strand as a template. A ribozyme on the 3' end of the gRNA cleaves the gRNA 3' of the second strand primer sequence. The newly synthesized double stranded DNA may be incorporated into the target nucleic acid at the site of the nick.

[0045] FIG. 4B shows the editing efficiency of a nCas9-RT construct using a pegRNA or a Stitch Guide gRNA. Schematics of the pegRNA and the Stitch Guide gRNA are shown at left. The fused nCas9-mlvRTv construct was used in this assay.

[0046] FIG. 5A shows the editing efficiency of a fused nCas9-RT construct ("nCas9-mlvRTv") with different gRNAs comprising second strand primers (SSPs) 20 nucleotides (nt), 40 nt, or 60 nt in length positioned either 6 nt, 36, nt, or 55 nt 3' of the 5' end of the first strand primer binding site ("nt from nick"). A gRNA lacking a second strand primer was tested as a control. All gRNA sequences comprised an HDV ribozyme (SEQ ID NO: 25).

[0047] FIG. 5B shows the editing efficiency of a nCas9-RT ("nCas9-R2(116-1016)") with different gRNAs comprising second strand primers (SSPs) 20 nucleotides (nt), 40 nt, or 60 nt in length that positioned either 6 nt, 36, nt, or 55 nt 3' of the 5' end of the first strand primer binding site ("nt from nick"). A gRNA lacking a second strand primer was tested as a control.

[0048] FIG. 6 illustrates four schemes of genome editing using a two gRNA system with a nCas9-RT. In a two single guide system in which the two guides each generate an edited strand (top left), each gRNA binds to a different nCas9 and the two gRNAs each comprise a reverse transcriptase template region. In a two single guide system in which the second guide nicks the opposite strand (top right), each gRNA binds to a different nCas9 and only one of the gRNAs comprise a reverse transcriptase template region. In a dual guide complex system in which the two guides each generate an edit (bottom left), the spacer of the first gRNA binds the first nCas9, the spacer of the second gRNA binds the second nCas9, the scaffold of the first gRNA binds the second nCas9, and the scaffold of the second gRNA binds the first nCas9; and the two gRNAs each comprise a reverse transcriptase template region and a primer binding site (PBS) region. In a dual guide complex system in which the second guide nicks the opposite strand (bottom right), the spacer of the first gRNA binds the first nCas9, the spacer of the second gRNA binds the second nCas9, the scaffold of the first gRNA binds the second nCas9, and the scaffold of the

second gRNA binds the first nCas9; and only one of the gRNAs comprise a reverse transcriptase template region.

[0049] FIG. 7 illustrates a method for increasing the efficiency of gene editing. A two single guide system in which the second guide nicks the opposite strand or a dual guide complex system in which the second guide nicks the opposite strand, the nick on the opposite strand facilitates incorporation of the newly synthesized DNA into the target nucleic acid. The second guide generates a flap that is reverse complementary to a region in the first newly synthesized strand. The first synthesized strand acts as template for second strand synthesis.

[0050] FIG. 8A illustrates a gRNA comprising a Velcro region to accelerate the rate of hybridization of the primer binding site and the flap by creating regions of reverse complementation within the 3' extended guide RNA. The Velcro region comprises 5 to 200 nucleotides positioned 5' of the reverse transcriptase template region that are reverse complementary to the region of the gRNA 5' of the first strand primer binding site.

[0051] FIG. 8B illustrates a gRNA comprising a Velcro region to accelerate the rate of hybridization of the primer binding site and the flap by creating regions of reverse complementation within the 3' extended guide RNA. The Velcro region comprises 5 to 100 nucleotides positioned 3' of the first strand primer binding site that are reverse complementary to a region of the reverse transcriptase template region.

[0052] FIG. 9A shows the editing efficiency of a nCas9-LZ1 and LZ2-mlvRTv construct with the gRNA constructs comprising a Velcro region, as illustrated in **FIG. 8A** and **FIG. 8B**. Editing efficiency was compared using a gRNA lacking a Velcro region ("no Velcro"), a 15 nt Velcro region positioned 5' of the reverse transcriptase template region ("V1," as illustrated in **FIG. 8A**) with a gap length of 1, 5, or 10 nts, or a Velcro region positioned 3' of the first strand primer binding site ("V2," as illustrated in **FIG. 8B**) of either 10 or 20 nt in length. The gRNA contained a 107 nucleotide RT template, and a 13 nucleotide primer binding site. Editing was performed such that an ATCC sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0053] FIG. 9B shows the editing efficiency of a nCas9-LZ1 and LZ2-R2(116-1016) construct with the gRNA constructs comprising a Velcro region, as illustrated in **FIG. 8A** and **FIG. 8B**. Editing efficiency was compared using a gRNA lacking a Velcro region ("no Velcro"), a 15 nt Velcro region positioned 5' of the reverse transcriptase template region ("V1," as illustrated in **FIG. 8A**) with a gap length of 1, 5, or 10 nts between the end of the Velcro binding site and the beginning of the primer binding site, or a Velcro region positioned 3' of the first strand primer binding site ("V2," as illustrated in **FIG. 8B**) of either 10 or 20 nt in length.

[0054] FIG. 10 shows the editing efficiency of a two gRNA system as illustrated in **FIG. 7**. A target nucleic acid encoding a blue fluorescent protein (BFP) was edited to introduce a stop codon. Lack of BFP fluorescence in a cell was indicative of successful editing. Editing efficiency, as measured by percent of cells negative for BFP (“% BFP-”), was measured for cells only (no gRNA), single gRNAs (gRNA 2 which lacks a 3’ extension, gRNA 1 without a stub, gRNA 1 with a stub), and two gRNAs (gRNA 1 without a stub plus gRNA 2 and gRNA 1 with a stub and gRNA 2).

[0055] FIG. 11A shows domain arrangements of a prime editor 2 system (“PE2,” top), a split prime editor 2 system (“split PE2,” middle), and a split stitch construct with two leucine zippers (“Split Stitch,” bottom). On the right is a structural schematic of the Split Stitch construct comprising a Cas9 nickase (nCas9) and a reverse transcriptase (RT) linked by two leucine zippers (LZ1 and LZ2) complexed with a guide nucleic acid. The Split Stitch split nCas9-LZ1 and LZ2-mlvRT construct comprises a nCas9-LZ1 (SEQ ID NO: 1, NLS-SpCas9(H840A)-NLS-EE12RR345L(leucine zipper)) and a LZ2-mlvRT (SEQ ID NO: 2, RR12EE345L(leucine zipper)-mlvRTv(nCas9-mlvRT(D200N, L603W, T306K, W313F, T330P)-NLS) on discrete polypeptide chains.

[0056] FIG. 11B shows the editing efficiency of the constructs illustrated in **FIG. 11A** with different gRNAs. Editing efficiency was measured as a percentage of cells that were edited to convert a BFP to a GFP (% GFP+). Editing efficiency was tested with different guide RNA (gRNA) constructs including gRNA 2.0 (SEQ ID NO: 31), a gRNA with a long MS2 hairpin (SEQ ID NO: 28), “gRNA-1xlongMS2”), a gRNA with a short MS2 hairpin (SEQ ID NO: 29, “gRNA-1xshortMS2”), or a gRNA with two short MS2 hairpins (SEQ ID NO: 30, “gRNA-2xshortMS2”).

[0057] FIG. 12A illustrates gRNA constructs either without (left) or with (middle and right) a Velcro region to accelerate the rate of hybridization of the primer binding site (PBS) to a flap of a target nucleic acid. In a V1 arrangement, the Velcro region may be positioned at or near the 5’ end of the gRNA and may hybridize to a region of the gRNA 5’ of the primer binding site (“Velcro V1,” middle). In a V2 arrangement, the Velcro region may be positioned 3’ of the primer binding site and may hybridize to a region at or near the 5’ end of the gRNA (“Velcro V2,” right).

[0058] FIG. 12B illustrates predicted three-dimensional structures of the gRNA constructs provided in **FIG. 12A**. A gRNA lacking a Velcro region is shown in the left. gRNAs comprising a Velcro V1 region or a Velcro V2 region are shown in the middle and right panels, respectively.

[0059] FIG. 12C shows editing efficiency of a gRNA with a 129 nucleotide RT template and a 13 nucleotide primer binding site and a 20 nucleotide Velcro region. Editing was performed such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT. Editing efficiency was compared for the original gRNA ("original coding") or a gRNA recoded with silent mutations in the RT template region of the gRNA extension to remove secondary structure ("recoded").

[0060] FIG. 12D shows editing efficiency of gRNAs with different lengths of Velcro sequences. Each gRNA contained, in order from 5' to 3', a RT template, a primer binding site, and a Velcro region, as shown in the schematic on the left. Editing efficiency was measured as the percent of cells that were GFP positive (% GFP+). gRNAs had a 129 nucleotide RT template, a 13 nucleotide primer binding site. Editing was performed such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT.

[0061] FIG. 13A illustrates schematics of a pegRNA and a Stitch gRNA comprising a Velcro region and a 2nd strand primer (top) and a method of genome editing using a Stitch gRNA (bottom). A nCas9-RT construct complexed with a gRNA is recruited to a target site of a target nucleic acid by hybridization of a spacer of the gRNA to the target site. The nCas9 nicks a strand of a target nucleic acid at a target site. A first strand primer binding site of the gRNA hybridizes to a flap 5' of the nick. The RT polymerizes from the 3' end of the flap using a reverse transcriptase template region of the gRNA as a template. A second strand primer ("2nd strand primer") at the 3' end of the gRNA hybridizes to the 3' end of the newly synthesized DNA strand. The 4-200 bp second strand primer region acts as an RNA primer for synthesis of a second DNA strand. The RT polymerizes from the 3' end of the gRNA using the newly synthesized DNA strand as a template. A ribozyme on the 3' end of the gRNA cleaves the gRNA 3' of the second strand primer sequence. The newly synthesized double stranded DNA may be incorporated into the target nucleic acid at the site of the nick.

[0062] FIG. 13B shows editing efficiency of gRNAs second strand primers (SSPs) of varying lengths and that hybridize at varying distances from the nicking site. Second strand primers 20, 40, or 60 nucleotides (nt) long positioned 6, 36, or 55 nucleotides from the nick were tested. Editing efficiency was measured as the percent of cells that were GFP positive (% GFP+).

[0063] FIG. 13C shows editing efficiency of gRNAs without a Velcro region or a second strand primer ("no velcro, no SSP"), with a 19 nucleotide Velcro region ("19 nt velcro"), or with both a 19 nucleotide Velcro region and a 20 nucleotide second strand primer ("19 nt velcro, 20 nt SSP").

[0064] FIG. 14A shows the results of a screen for mutations in a mlvRT reverse transcriptase and their effect on editing efficiency. Mutations were made in a reference mlvRT construct containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing a Y8H, P51L, S56A, S67R, E69K, Q84A, F155Y, T197A, H204R, T246E, N249D, E286R, Q291I, R301L, E302K, F309N, M320L, L435G, D524A, D524G, D524N, E562D, K571R, D583N, Y586S, H594Q, H638G, D653N, T664N, or L671P single point mutation (SEQ ID NO: 41 – SEQ ID NO: 70, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 1 nucleotide gap, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA. Mutation rate data are shown as mean $\pm$ one standard deviation from three biologically independent samples.

[0065] FIG. 14B shows the results of a screen for combinations of mutations in a mlvRT reverse transcriptase and their effect on editing efficiency. Mutations were made in a reference mlvRT construct containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing T197A and D653N; T197A and T664N; T197A and L671P; T197A, D653N, T664N and L671P; or P51L, S67R, T197A, H204R, L435G, D524A, D653N, T664N and L671P (SEQ ID NO: 71 – SEQ ID NO: 75, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0066] FIG. 15A illustrates a method of increasing availability of dNTPs in a cell to increase editing efficiency. In non-dividing cells lacking CDK1, unphosphorylated SAMHD1 cleaves dNTPs, decreasing the available dNTPs in the cell. In dividing cells, CDK1 phosphorylates SAMHD1, preventing SAMHD1 from cleaving dNTPs and leading to increased availability of dNTPs in the cell. A single point mutation in SAMHD1 (T592A) prevents phosphorylation of SAMHD1 by CDK1, resulting in a constitutively active SAMHD1 and a low availability of dNTPs in the cell. The T592A mutant SAMHD1 was used to induce a low dNTP environment in

the assay shown in **FIG. 15B**, **FIG. 15D**, and **FIG. 15E**. Addition of Vpx inhibits SAMHD1, leading to increased availability of dNTPs in the cell.

[0067] FIG. 15B shows the editing efficiency of mlvRT reverse transcriptase constructs in the presence or absence of a constitutively active SAMHD1 (SAMHD1 (T592A)) to decrease availability of dNTPs in the cell. Mutations were made in a reference mlvRT construct containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing Q221R; V223A; V223M; Q221R and V223A; or Q221R and V223M (SEQ ID NO: 76 – SEQ ID NO: 80, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0068] FIG. 15C shows the editing efficiency of mlvRT reverse transcriptase constructs. Mutations were made in a reference mlvRT construct containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing V223A; V223M; Q221R and V223A; or Q221R and V223M (SEQ ID NO: 77 – SEQ ID NO: 80, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with a 129 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT.

[0069] FIG. 15D shows the editing efficiency of a mlvRT reverse transcriptase in the presence or absence of a constitutively active SAMHD1 (SAMHD1 (T592A)) to decrease availability of dNTPs in the cell and with or without Vpx (SEQ ID NO: 82) to inhibit SAMHD1. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0070] FIG. 15E shows the editing efficiency of a mlvRT reverse transcriptase in the presence or absence of a constitutively active SAMHD1 (SAMHD1 (T592A)) to decrease availability of dNTPs in the cell and with or without Vpx (SEQ ID NO: 82) to inhibit SAMHD1. Editing

efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with a 129 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT.

[0071] FIG. 15F shows that coexpression of VPX^{RH2-1} and SAMHD1 (T592A) with RW2I completely reversed the reduction in editing efficiency caused by expressing SAMHD1 (T592A) without VPX.

[0072] FIG. 16 shows editing efficiency of Cas9 constructs modified for nicking activity and linked to a reverse transcriptase through a leucine zipper. *S. Pyogenes* Cas9 ("SpCas9") constructs contained an H840A mutation to produce a Cas9 nickase (nCas9). Cysteine residues were introduced into the Cas9 nickase at either D1079C, S1173C, or D1180C to enable splitting of the Cas9 into a split intein Cas9 (iCas9) for expression as extein-intein fusions. Leucine zipper Cas9 constructs containing H840A and D1079C (SEQ ID NO: 85 with a leucine zipper), H840A and S1173C (SEQ ID NO: 86 with a leucine zipper), or H840A and D1180C (SEQ ID NO: 87 with a leucine zipper) point mutations and linked to mlvRT5M (SEQ ID NO: 40 with a leucine zipper) were tested. A Cas9 nickase that contained the H840A mutation but no additional cysteine (SEQ ID NO: 84 with a leucine zipper) linked to mlvRT5M (SEQ ID NO: 40 with a leucine zipper) was used as a control. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0073] FIG. 17A shows the editing efficiency of a split intein Cas9 (iCas9) S1173C construct modified for nicking activity, fused to a reverse transcriptase, and expressed as two extein-intein fusion proteins. The N-terminal region of the nCas9-RT construct was expressed as nCas9(1-1172) – Npu N intein (SEQ ID NO: 90) and the C-terminal region of the nCas9-RT construct was expressed as Npu C intein – nCas9(1173-1368 with S1173C) – mlvRT5M (SEQ ID NO: 91). Editing efficiency of the split intein Cas9-RT construct (right bar) was compared to a leucine zipper split Cas9 construct (SEQ ID NO: 1 and SEQ ID NO: 2, left bar). Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0074] FIG. 17B shows the editing efficiency of a split intein Cas9 (iCas9) S1173C construct modified for nicking activity, fused to a reverse transcriptase, and expressed as two extein-intein fusion proteins. The N-terminal region of the nCas9-RT construct was expressed as nCas9(1-1172) – Npu N intein (SEQ ID NO: 90) and the C-terminal region of the nCas9-RT construct was expressed as Npu C intein – nCas9(1173-1368 with S1173C) – mlvRT5M (SEQ ID NO: 91). Editing efficiency of the split intein Cas9-RT construct (right bar) was compared to a leucine zipper split Cas9 construct (SEQ ID NO: 1 and SEQ ID NO: 2, left bar). Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT.

[0075] FIG. 18 shows the editing efficiency of a leucine zipper Cas9-RT construct in the presence of a gRNA comprising either an HDV ribozyme (left bar) or a tRNA (right bar) at the 3' end of the gRNA, immediately 3' of the second strand primer. The leucine zipper Cas9-RT construct was expressed as nCas9-LZ1 (SEQ ID NO: 1) and LZ2-mlvRT5M (SEQ ID NO: 2) and linked through a leucine zipper. The tRNA had a sequence corresponding to SEQ ID NO: 94 (GGTCCCATGGTGTAATGGTTAGCACTCTGGACTTTGAATCCAGCGATCCGAGTTCA AATCTCGGTGGGACCT). Editing was performed using gRNAs with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, a 20 nucleotide second strand primer, and either an HDV ribozyme or a tRNA 3' of the second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0076] FIG. 19A shows abilities of some precision editing constructs to fit in an AAV, or not to fit within AAV. Both Prime Editor 2 (PE2) and Split PE2 may utilize a nicking Cas9 (nCas9) fused to a Moloney leukemia virus reverse transcriptase pentamutant (mlvRT5M), which may be encoded by ORFs too large to be packaged into AAV. Architectures were developed that can deliver nCas9 and mlvRT5M with AAV, each encoded by two ORFs that are each smaller than the carrying capacity of AAV. Rewriter a1 (RWa1) may utilize nCas9, a MS2 coat protein (MCP) peptide fused to mlvRT5M and a gRNA comprising the MS2 hairpin to which the MCP peptide may specifically bind. RWb1 may utilize heterodimerizing leucine zippers to colocalize nCas9 and mlvRT5M. RWc1 may utilize a novel mutant nCas9 that can be split with Npu inteins to produce a nCas9-mlvRT5M protein using ORFs that each fit in AAV. RWd1 may coexpress nCas9 and mlvRT5M without any engineered recruitment components.

[0077] FIG. 19B shows GFP expression in cells comprising some editing constructs. Testing a panel of N-terminal fragments of nCas9 fused to the Npu N-terminal intein paired with an Npu C-terminal intein fused to a C-terminal fragment of nCas9 comprising a cysteine substitution providing for intein catalysis and mlvRT5M revealed that splitting a Ser1173Cys nCas9-mlvRT5M mutant between residues 1172 and 1173 (nCas9(1-1172)-NpuN and nCas9(1173-1368; S1173C)-mlvRT5M) resulted in at least about a 2-fold greater editing efficiency than PE2.

[0078] FIG. 19C shows GFP expression in cells comprising some editing constructs.

Cotransfecting RWa1, RWb1, and RWc1 with the standard gRNA scaffold all resulted in above 40% editing efficiency. The editing efficiency over 40% was achieved with both RW1M paired with a RTT-PBS-MS2 gRNA extension architecture and RW1L achieving over 50% editing.

Coexpression of a gRNA that does not contain the MS2 hairpins with the nCas9 and MCP-mlvRT5M constructs resulted in approximately the same editing efficiency compared to gRNAs that included an MS2 hairpin (RWd1).

[0079] FIG. 19D shows GFP expression in cells transduced with some editing constructs.

Transducing HEK293 cells expressing BFP with RWc1 packaged into two separate AAV2 constructs at an MOI of 2.8×10^5 for each virus resulted in 74.8% GFP+ cells. Mutation rate data are shown as mean $\pm$ one standard deviation from three biologically independent samples.

[0080] FIG. 20A shows some spatial orientations of enzymes and guide RNAs. Reverse transcription of the RTT can in some cases only be initiated after the PBS hybridizes to the genomic flap. Inserting either a sequence 5' of the RTT that hybridizes with a 3' region of the RTT (GPS V1) or a sequence 3' of the PBS that hybridizes with the 5' portion of the RTT (GPS V2) may reorient the PBS to be in closer proximity to the genomic flap.

[0081] FIG. 20B shows GFP expression in cells comprising some editing constructs. RWb1 and a guide RNA comprising a 107-nt RTT resulted in 14% GFP+ cells, which is significantly lower than the 38% achieved using a shorter 13-nt RTT. Adding a 20-nt GPS V2 (RWb2) increased the editing efficiency to ~27%.

[0082] FIG. 20C shows GFP expression in cells comprising some editing constructs. Installing a 3-nt mutation 65-nt from the site of the nick using a 129-nt template was increased 4-fold by incorporating GPS V2.

[0083] FIG. 21 shows GFP expression in cells comprising some editing constructs. Velcro and SSP may be used simultaneously, resulting in ~41% editing (Rewriter 3.2). The increase in efficiency that SSP provided was abolished when the terminal 3-nt of SSP were not complementary to the first synthesized strand. Mutation rate data are shown as mean $\pm$ one

standard deviation from three biologically independent samples. * = $P < 0.05$; two-sided student's t-test.

[0084] **FIG. 22** shows that incorporating a human glutamate tRNA after SSP led to a statistically significant increase in editing efficiency compared to an HDV ribozyme following SSP.

Mutation rate data are shown as mean $\pm$ one standard deviation from three biologically independent samples. * = $P < 0.05$; two-sided student's t-test.

[0085] **FIG. 23** shows that Coexpression of SAMHD1p- with Rewriter 3.0 drastically decreased the efficiency of installing a mutation 65-nt from the nick. Additional coexpression of VPXROD restored the editing efficiency to 78% of the efficiency observed in the absence of SAMHD1p-.

Mutation rate data are mean $\pm$ one standard deviation from three biologically independent samples.

[0086] **FIG. 24** is a graph showing editing efficiencies of various editing components expressed together in cells.

[0087] **FIG. 25A** is a chart showing editing efficiencies using Rewriter constructs.

[0088] **FIGs. 25B and 25C** illustrate information about some experiments performed using guide RNAs.

[0089] **FIG. 25D** illustrates that transfecting HEK293T with RW2I and a gRNA to install the 2298T>C mutation did not introduce mutations at the spacer's top five in silico-predicted off-target sites.

[0090] **FIG. 25E** illustrates that modifying the RTT to include a silent 2307A>G mutation that disrupts a polyA track eliminated an undesirable insertion of an adenine.

[0091] **FIG. 25F** illustrates that screening additional RTT lengths increased editing efficiency to 41.6%.

[0092] **FIG. 25G** illustrates that encoding a silent mutation in the RTT that would disrupt the spacer's PAM sequence doubled the efficiency of installing the 2298T>C mutation.

[0093] **FIG. 26** shows exemplary configurations of Velcro (also referred to as GPS) in a guide nucleic acid.

[0094] **FIG. 27** shows components of some editing systems.

[0095] **FIG. 28** shows editing efficiencies obtained using some editing system components.

[0096] **FIGs. 29A-29F** show % reads of nucleobases after use of some editing systems.

[0097] **FIG. 30** shows components of some editing systems.

[0098] **FIG. 31A** shows % of reads with mutations after treatment with some editing components described herein.

[0099] **FIG. 31B** shows % of reads with mutations in wildtype cells.

[0100] FIG. 32 shows a dual guide system.

[0101] FIG. 33 shows editing efficiencies with a dual guide system.

DETAILED DESCRIPTION

[0102] Disclosed herein are methods and compositions for precise and efficient genome editing using CRISPR-Cas systems. Cas9-based base editors comprising a Cas9 nickase (nCas9) linked to a deaminase may be limited to performing transition mutations (e.g., A to G or C to T). Other Cas9-based editors (e.g., “prime editors”) comprising a nCas9 linked to a reverse transcriptase (RT) (e.g., a Moloney leukemia virus RT) may be limited to small insertions, deletions, or single nucleotide changes. Provided herein are Cas nickase and reverse transcriptase constructs, engineered guide nucleic acids, and methods of use thereof for improved efficiency, versatility, precision, and deliverability of genome editing.

[0103] The methods and compositions described herein may include splitting, dimerizing, or coexpressing a Cas9 and a RT. The splitting, dimerizing, or coexpressing of Cas9 and RT may enable AAV packaging. The splitting, dimerizing, or coexpressing of Cas9 and RT may increase editing efficiency.

[0104] Described herein are AAV deliverable precision editing components. Some embodiments include AAV particles that deliver a Cas9 component and a RT component. Various examples are provided for delivering Cas+RT systems with AAV. The examples provided may overcome previous difficulties getting precision editing components to fit within a typical AAV carrying capacity (e.g. of about 4.5 kb).

[0105] Also provided are mutations such as point mutations or insertion mutations that improve editing efficiency. For example, Cas nickase or RT (e.g. point mutations or insertion mutations) are included. Some embodiments include an mlvRT for genome editing with an amino acid mutation.

Nicking Cas9 and Reverse Transcriptase Enzymes

[0106] Provided herein are compositions comprising a Cas nickase. Provided herein are compositions comprising a reverse transcriptase. Provided herein are compositions comprising a Cas nickase and a reverse transcriptase. At least part of the Cas nickase and the reverse transcriptase may be included in separate polypeptide chains. The Cas nickase and the reverse transcriptase may be completely in separate polypeptide chains. Some embodiments include a functional fragment of the Cas nickase. Some embodiments include a functional fragment of the reverse transcriptase.

[0107] The Cas nickase and the reverse transcriptase may form a Cas-reverse transcriptase heterodimer. The Cas-reverse transcriptase heterodimer may include a first heterodimer domain. The first heterodimer domain may be fused to the Cas nickase. The Cas-reverse transcriptase heterodimer may include a second heterodimer domain. The second heterodimer domain may be fused to the reverse transcriptase. The first heterodimer domain may bind the second heterodimer domain. This binding may form the Cas-reverse transcriptase heterodimer. The first heterodimer domain may comprise a leucine zipper. The second heterodimer domain may comprise a leucine zipper. The first or second heterodimer domain may comprise a heterodimer domain other than a leucine zipper, for example a SpyCatcher or SpyTag moiety as described herein.

[0108] Provided herein are engineered constructs comprising a Cas programmable nuclease. The Cas programmable nuclease may comprise a Cas9 programmable nuclease. Provided herein are engineered constructs comprising a Cas nickase. The Cas programmable nuclease may include a Cas nickase. The Cas nickase may comprise a Cas9 nickase (nCas9). The Cas9 programmable nuclease may comprise an nCas9. The Cas nickase may be generated by mutating a Cas9 nuclease domain. The Cas nickase may create a single-strand rather than a double-strand break.

[0109] Provided herein are engineered constructs comprising a reverse transcriptase (RT). Provided herein are engineered constructs comprising a Cas nickase and a RT. Provided herein are engineered constructs comprising a Cas9 nickase and a RT. The nCas9 may introduce a single-strand break (SSB) at a target site of a target nucleic acid. The reverse transcriptase may catalyze reverse transcription of a sequence to be inserted at the target site. In some embodiments, a nCas9-RT construct may be fused to a nCas9-RT construct. A fused nCas9-RT construct may comprise a nCas9 and a reverse transcriptase in a single polypeptide chain. In some embodiments, a nCas9-RT construct may be a split nCas9-RT construct. A split nCas9-RT construct may comprise a nCas9 in a first polypeptide chain and a reverse transcriptase in a second polypeptide chain. The nCas9 and the reverse transcriptase of a split nCas9-RT construct may form a heterodimer when co-expressed. In some embodiments, a first dimerization domain may be located N-terminal of the nCas9. In some embodiments, a second dimerization domain that dimerizes with the first dimerization domain may be located C-terminal of the reverse transcriptase. In some embodiments, a first dimerization domain may be located C-terminal of the nCas9. In some embodiments, a second dimerization domain that dimerizes with the first dimerization domain may be located N-terminal of the reverse transcriptase. The first dimerization domain may comprise a leucine zipper, an FKBP, an FRB, a Calcineurin A, a CyP-Fas, a GyrB, a GAI, a GID1, a SNAP tag, a Halo tag, a Bcl-xL, a Fab, or a LOV domain. The second dimerization domain may comprise a leucine zipper, an FKBP, an FRB, a Calcineurin A,

a CyP-Fas, a GyrB, a GAI, a GID1, a SNAP tag, a Halo tag, a Bcl-xL, a Fab, or a LOV domain. Dimerization may be induced or spontaneous. Dimerization may be chemically or optically induced. SEQ ID NO:1 provides an example of a nCas9 comprising a leucine zipper at the C-terminus. SEQ ID NO: 2 provides an example of a reverse transcriptase comprising a leucine zipper at the N-terminus.

[0110] In some embodiments, a construct of the present disclosure may comprise a nuclear localization signal (NLS). A composition described herein may comprise a nuclear localization signal fused to a Cas nickase. In some embodiments, the Cas nickase fused to an NLS comprises a sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 138. A composition described herein may comprise a nuclear localization signal fused to a RT. In some embodiments, the RT fused to an NLS comprises a sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 95.

[0111] The reverse transcriptase may comprise a domain from a non-long terminal repeat retrotransposable element. The non-long terminal repeat retrotransposable element may be fused to part of a Cas nickase. The reverse transcriptase may comprise a sequence from a bacterial group II intron. The bacterial group II intron may be fused to part of the Cas nickase. The reverse transcriptase may comprise a domain from a retroviral gag-pol polyprotein. The domain from the retroviral gag-pol polyprotein may be fused to part of the Cas nickase.

[0112] Dimerization may be achieved using a SpyTag/SpyCatcher or related system. For example, a RT may be conjugated to a SpyTag moiety, and a Cas nickase may be conjugated to a SpyCatcher moiety. Alternatively, a Cas nickase may be conjugated to a SpyTag moiety, and a RT may be conjugated to a SpyCatcher moiety. Dimerization using the SpyTag/SpyCatcher system may include covalent bonds between dimerized molecules (e.g. the Cas nickase may be covalently conjugated to the RT through the SpyTag and SpyCatcher moieties. A Cas nickase conjugated to a SpyTag or SpyCatcher moiety may be provided in a first AAV. A RT conjugated to a SpyCatcher or SpyTag moiety may be provided in a second AAV.

[0113] A variety of reverse transcriptases are consistent with the compositions and methods of the present disclosure. A reverse transcriptase as disclosed herein may be a *geobacillus stereothermophilus* RT (GsI-IICRT, SEQ ID NO: 3), *Eubacterium Rectale* RT (ErRT, SEQ ID NO: 4), marathon RT (SEQ ID NO: 5), BmR2RT (SEQ ID NO: 6), amino acids 116-1016 from the R2 polyprotein (R2(116-1016), SEQ ID NO: 7), BmR2en-RT (SEQ ID NO: 8), humanL1RT (SEQ ID NO: 9), humanL1en-RT (SEQ ID NO: 10), murineL1RT (SEQ ID NO: 11), ltrA (SEQ

ID NO: 12), mlvRT5M (SEQ ID NO: 13), mlvRT5M (SEQ ID NO: 40), mlvRT (SEQ ID NO: 14), XMRV3VP35RT (SEQ ID NO: 15), galvRT (SEQ ID NO: 16), sfvRT (SEQ ID NO: 17), foamvRT (SEQ ID NO: 18), HIVP66 (SEQ ID NO: 19), HIVP51 (SEQ ID NO: 20), rsvAlpha (SEQ ID NO: 21), or rsvBeta (SEQ ID NO: 22). A transcriptase of the present disclosure may include an N-terminal methionine, or a transcriptase of the present disclosure may lack an N-terminal methionine. For example, a reverse transcriptase may have a sequence corresponding to any one of SEQ ID NO: 3 – SEQ ID NO: 6, SEQ ID NO: 8 – SEQ ID NO: 12, SEQ ID NO: 17, SEQ ID NO: 18, or SEQ ID NO: 40 – SEQ ID NO: 80 with the N-terminal methionine removed. In another example, a reverse transcriptase may have a sequence corresponding to any one of SEQ ID NO: 7, SEQ ID NO: 13 – SEQ ID NO: 16, or SEQ ID NO: 19 – SEQ ID NO: 22 with a methionine added to the N-terminus. A reverse transcriptase may comprise a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, 100% sequence identity, or any percentage therebetween, to any one of SEQ ID NO: 3 – SEQ ID NO: 22 or SEQ ID NO: 40 – SEQ ID NO: 80, or a fragment thereof.

[0114] Disclosed herein are compositions comprising a reverse transcriptase with a sequence having at least 80% sequence identity to of any one of SEQ ID NO: 3 – SEQ ID NO: 22 or SEQ ID NO: 40 – SEQ ID NO: 80, a fragment thereof. The reverse transcriptase or fragment thereof may be fused to a Cas nickase. Some embodiments include a composition comprising a reverse transcriptase comprising a domain from a non-long terminal repeat retrotransposable element, which may be fused to a Cas nickase. Some embodiments include a reverse transcriptase comprising a sequence from a bacterial group II intron, which may be fused to a Cas nickase. Some embodiments include a reverse transcriptase comprising a domain from a retroviral gag-pol polyprotein, that may be fused to a Cas nickase. The reverse transcriptase may be truncated.

[0115] Disclosed are methods of optimizing genome editing efficiency in a limiting dNTP condition. The method may include performing genome editing with a Moloney leukemia virus reverse transcriptase (mlvRT), or a variant thereof. A RT described herein such as a mlvRT may include a mutation such as a point mutation. The point mutation may be at position 221 of the reverse transcriptase. The mlvRT or variant thereof may comprise a point mutation at position 221. The point mutation at position 221 may comprise Q221R. The point mutation may be at position 223 of the reverse transcriptase. The mlvRT or variant thereof may comprise a point mutation at position 223. The point mutation at position 223 may comprise V223A. The point mutation at position 223 may comprise V223M.

[0116] Some embodiments include a method of optimizing genome editing efficiency, comprising performing genome editing with a Moloney leukemia virus reverse transcriptase

(mlvRT) that is modified to increase its catalytic efficiency in low dNTP concentrations. For example, the mlvRT may be modified to decrease its K_m for dNTPs.

[0117] A reverse transcriptase of the present disclosure may comprise one or more mutations. For example, a reverse transcriptase may comprise one or more mutations relative to a reference reverse transcriptase sequence (e.g., SEQ ID NO: 81). In some embodiments, a point mutation in a reverse transcriptase may increase the editing efficiency of a Cas9-RT construct relative to a reference sequence lacking the point mutation. A reverse transcriptase may comprise one or more mutations corresponding to D200N, I603W, T330P, T306K, W313F, Y8H, P51L, S56A, S67R, E69K, Q84A, F155Y, T197A, H204R, T246E, N249D, E286R, Q291I, R301L, E302K, F309N, M320L, L435G, D524A, D524G, D524N, E562D, K571R, D583N, Y586S, H594Q, H638G, D653N, T664N, L671P, Q221R, V223A, V223M, or combinations thereof, relative to SEQ ID NO: 81. A reverse transcriptase may comprise one or more mutations (e.g. point mutations) at amino acid position Q84, L139, Q221, V223, T664, L671, D524, P51, or S67. A reverse transcriptase may comprise one or more mutations (e.g. point mutations) corresponding to Q84A, L139P, Q221R, V223A, V223M, T664N, L671P, D524A, P51L, or S67R. The one or more mutations may be in relation to SEQ ID NO: 81 or another sequence identified herein. The one or more mutations may be in relation to an amino acid sequence at least 75%, identical at least 80%, identical at least 85%, identical at least 86%, identical at least 87%, identical at least 88%, identical at least 89%, identical at least 90%, identical at least 91%, identical at least 92%, identical at least 93%, identical at least 94%, identical at least 95%, identical at least 96%, identical at least 97%, identical at least 98%, identical or at least 99% identical, to SEQ ID NO: 81 or another sequence identified herein. In some embodiments, a reverse transcriptase may comprise mutations corresponding to D200N, I603W, T330P, T306K, and W313F (e.g., SEQ ID NO: 13 or SEQ ID NO: 40). In some embodiments, a reverse transcriptase may comprise mutations corresponding to D200N, I603W, T330P, T306K, and W313F and one or more additional mutations (e.g., SEQ ID NO: 41 – SEQ ID NO: 80).

[0118] The RT may include one or more mutations included in **FIG. 14B**. For example, the RT may include a mutation at position 51, 67, 84, 139, 197, 204, 435, 524, 653, 664, or 671, or a combination thereof. The RT may include a mutation at position P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524, or a combination thereof. The RT may include mutations at position 51, 67, 84, 139, 197, 204, 435, 524, 653, 664, and 671. The mutation may include at least one point mutation. The at least one point mutation may be at P51L, S67R, Q84A, L139P, Q221R, V223A, V223M, T197A, D653N, T664N, L671P, L435G, H204R, or D524A, or a combination thereof. The RT may include a mutation at position 51. The

mutation at position 51 may include P51L. The RT may include a mutation at position 67. The mutation at position 67 may include S67R. The RT may include a mutation at position 84. The mutation at position 84 may include Q84A. The RT may include a mutation at position 139. The mutation at position 139 may include L139P. The RT may include a mutation at position 197. The mutation at position 197 may include T197A. The RT may include a mutation at position 204. The mutation at position 204 may include H204R. The RT may include a mutation at position 435. The mutation at position 435 may include L435G. The RT may include a mutation at position 524. The mutation at position 524 may include D524A. The RT may include a mutation at position 653. The mutation at position 653 may include D653N. The RT may include a mutation at position 664. The mutation at position 664 may include T664N. The RT may include a mutation at position 671. The mutation at position 671 may include L671P. The RT may include, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more, of said mutations. The RT may include a mlvRT. The RT may include a mlvRT5M. The RT with the one or more mutations may comprise the one or more mutations with reference to a RT sequence provided herein. The one or more mutations may increase editing efficiency of a composition described herein, in relation to a composition without the one or more mutations. The mutation may be or include an insertion mutation. The reverse transcriptase comprises an insertion mutation immediately upstream (e.g. in the amino end direction) of P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524, or a combination thereof. The reverse transcriptase comprises an insertion mutation immediately downstream (e.g. in the carboxy end direction) of P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524, or a combination thereof. The insertion mutation may comprise an insertion of an amino acid disclosed herein for a point mutation, wherein the point mutation is to an amino acid.

[0119] In some embodiments, the reverse transcriptase has a point mutation at position P51, S67, Q84, L139, T197, D200, H204, Q221, V223, T306, W313, T330, L435, D524, D653, T664, L671, or L600, or a combination thereof. In some embodiments, the reverse transcriptase has a point mutation at position P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524, or a combination thereof. In some embodiments, the reverse transcriptase has a point mutation at position Q84, L139, Q221, V223, T664, or L671, or a combination thereof.

[0120] In some embodiments, the reverse transcriptase has a point mutation comprising P51L, S67R, Q84A, L139P, T197A, D200N, H204R, Q221R, V223A, V223M, T306K, W313F, T330P, L435G, D524A, D653N, T664N, L671P, or L603W, or a combination thereof. In some embodiments, the reverse transcriptase has a point mutation comprising P51L, S67R, Q84A, L139P, Q221R, V223A, V223M, T197A, D653N, T664N, L671P, L435G, H204R, or D524A, or

a combination thereof. In some embodiments, the reverse transcriptase has a point mutation comprising S67R, Q84A, L139P, Q221R, V223A, V223M, T664N, L671P, or D524A, or a combination thereof.

[0121] A reverse transcriptase of the present disclosure may comprise a loop region (e.g., a 2a loop or a 3a loop). A reverse transcriptase of the present disclosure may transcribe an editing sequence of 20 or more, 40 or more, 45 or more, 50 or more, 60 or more, 81 or more, 100 or more, 500 or more, 1000 or more, 2000 or more, 3000 or more, 4000 or more, 5000 or more, 6000 or more, 7000 or more, 8000 or more, 9000 or more, or 10,000 or more nucleotides. A reverse transcriptase of the present disclosure may transcribe an editing sequence of up to about 20, up to about 40, up to about 45, up to about 50, up to about 60, up to about 81, up to about 100, up to about 500, up to about 1000, up to about 2000, up to about 3000, up to about 4000, up to about 5000, up to about 6000, up to about 7000, up to about 8000, up to about 9000, or up to about 10,000 nucleotides. A reverse transcriptase of the present disclosure may transcribe an editing sequence of from 20 to 10,000 nucleotides.

[0122] A reverse transcriptase of the present disclosure can have increased processivity. Processivity may be determined by the number of phosphodiester bonds catalyzed by the reverse transcriptase in a single binding event. The processivity may be compared to a native reverse transcriptase. The reverse transcriptase may comprise increased processivity compared to a mlvRT. A reverse transcriptase with increased processivity may edit longer sequences at a target site of a target nucleic acid. For example, a reverse transcriptase with increased processivity may increase the editing window length of a programmable nuclease. The reverse transcriptase may edit a longer window length in a target sequence compared to a mlvRT. A reverse transcriptase with increased processivity may comprise an insert sequence. In some embodiments, an insertion that increases processivity may be inserted into a reverse transcriptase between domains 2 and 3 or between domains 3 and 4. A reverse transcriptase with increased processivity may comprise a deletion. For example, a reverse transcriptase with increased processivity may lack an RNase domain or may lack a connect domain. A reverse transcriptase with increased processivity may catalyze 20 or more, 40 or more, 45 or more, 50 or more, 60 or more, 81 or more, 100 or more, 500 or more, 1000 or more, 2000 or more, 3000 or more, 4000 or more, 5000 or more, 6000 or more, 7000 or more, 8000 or more, 9000 or more, or 10,000 or more phosphodiester bonds in a single binding event. A reverse transcriptase with increased processivity may catalyze up to about 20, up to about 40, up to about 45, up to about 50, up to about 60, up to about 81, up to about 100, up to about 500, up to about 1000, up to about 2000, up to about 3000, up to about

4000, up to about 5000, up to about 6000, up to about 7000, up to about 8000, up to about 9000, or up to about 10,000 phosphodiester bonds in a single binding event.

[0123] In some embodiments, a reverse transcriptase edits a longer sequence at a target site of a target nucleic acid than mlvRT. The reverse transcriptase may increase the editing window length of a programmable nuclease. A reverse transcriptase that edits a longer sequence at a target site may comprise an insert sequence. In some embodiments, an insertion is inserted into a reverse transcriptase that edits a longer sequence at a target site between domains 2 and 3 or between domains 3 and 4. A reverse transcriptase that edits a longer sequence at a target site may comprise a deletion. For example, a reverse transcriptase that edits a longer sequence at a target site may lack an RNase domain or may lack a connect domain. A reverse transcriptase that edits a longer sequence at a target site may catalyze 20 or more, 40 or more, 45 or more, 50 or more, 60 or more, 81 or more, 100 or more, 500 or more, 1000 or more, 2000 or more, 3000 or more, 4000 or more, 5000 or more, 6000 or more, 7000 or more, 8000 or more, 9000 or more, or 10,000 or more phosphodiester bonds in a single binding event. A reverse transcriptase that edits a longer sequence at a target site may catalyze up to about 20, up to about 40, up to about 45, up to about 50, up to about 60, up to about 81, up to about 100, up to about 500, up to about 1000, up to about 2000, up to about 3000, up to about 4000, up to about 5000, up to about 6000, up to about 7000, up to about 8000, up to about 9000, or up to about 10,000 phosphodiester bonds in a single binding event. In some embodiments, the reverse transcriptase that edits a longer sequence at a target site also has increased processivity as described herein.

[0124] A reverse transcriptase of the present disclosure may be a small reverse transcriptase. A small reverse transcriptase may have improved delivery to a cell as compared to a larger reverse transcriptase. The reverse transcriptase may comprise improved delivery to a cell compared to a mlvRT. A small reverse transcriptase may have improved expression in a cell as compared to a larger reverse transcriptase. A small reverse transcriptase may comprise no more than about 400, no more than about 420, no more than about 427, no more than about 440, no more than about 450, no more than about 500, no more than about 550, no more than about 560, no more than about 599, no more than about 600, no more than about 650, no more than about 677, no more than about 682, no more than about 700, no more than about 750, no more than about 761, no more than about 762, no more than about 800, no more than about 850, no more than about 900, no more than about 901, no more than about 950, no more than about 1000, no more than about 1100, no more than about 1114, no more than about 1200, no more than about 1275, no more than about 1281, or no more than about 1300 amino acid residues. A construct of the present disclosure may comprise a small reverse transcriptase, a dimerization region, a localization

region, or a combination thereof. A small reverse transcriptase may have increased processivity, edit a longer sequence at a target site, or a combination thereof.

[0125] A reverse transcriptase of the present disclosure may have a decreased immunogenicity as compared to a Moloney leukemia virus reverse transcriptase. A reverse transcriptase with decreased immunogenicity may also be a small reverse transcriptase, may have increased processivity, edit a longer sequence at a target site, or any combination thereof.

[0126] Disclosed herein are compositions comprising a Cas nickase or a Cas9 programmable nuclease. Examples of Cas nickases or Cas9 programmable nucleases that are consistent with the present disclosure include SpCas9 (SEQ ID NO: 32), SaCas9 (SEQ ID NO: 33), CjCas9 (SEQ ID NO: 34), GeoCas9 (SEQ ID NO: 35), HpaCas9 (SEQ ID NO: 36), and NmeCas9 (SEQ ID NO: 37). In some embodiments, the Cas nickase comprises a sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of SEQ ID NOs: 32-37. A Cas9 programmable nuclease of the present disclosure may comprise mutations, insertions, deletions, or truncations with respect to a native Cas9 programmable nuclease.

[0127] The Cas nickase may include a mutation. The mutation may enable or improve intein catalysis. The mutation may be an insertion mutation. The mutation may be a point mutation. The Cas nickase may include a cysteine point mutation. The cysteine point mutation may be located in a C-terminal half of the Cas nickase. A Cas9 described herein may include a cysteine point mutation in a C-terminal half of the Cas9. The cysteine point mutation may be located anywhere after amino acid position 574 of the Cas nickase. The mutation may be in an *S. Pyogenes* Cas9 nickase. The cysteine point mutation may comprise S1173. The cysteine point mutation may comprise D1079. The cysteine point mutation may comprise D1180.

[0128] The Cas9 nickase (an *S. Pyogenes* Cas9 nickase) may include a point mutation. The point mutation may enable intein catalysis. The point mutation may improve intein catalysis. In some embodiments, the point mutation comprises a cysteine point mutation, a serine point mutation, a threonine point mutation, or an alanine point mutation. In some embodiments, the point mutation comprises a cysteine point mutation. In some embodiments, the point mutation comprises a serine point mutation. In some embodiments, the point mutation comprises a threonine point mutation. In some embodiments, the point mutation comprises an alanine point mutation. In some embodiments, the point mutation is located at D1079. In some embodiments, the point mutation is located at D1125. In some embodiments, the point mutation is located at D1130. In some embodiments, the point mutation is located at G1133. In some embodiments, the point mutation is located at A1140. In some embodiments, the point mutation is located at I1168. In

some embodiments, the point mutation is located at S1173. In some embodiments, the point mutation is located at D1180. In some embodiments, the point mutation is located at G1186. In some embodiments, the point mutation is located at L1203. In some embodiments, the point mutation is located at R1212. In some embodiments, the point mutation is located at D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212 of the *S. Pyogenes* Cas9.

[0129] The Cas9 nickase (an *S. Pyogenes* Cas9 nickase) may include a insertion mutation. The insertion mutation may enable intein catalysis. The insertion mutation may improve intein catalysis. In some embodiments, the insertion mutation comprises a cysteine insertion mutation, a serine insertion mutation, a threonine insertion mutation, or an alanine insertion mutation. In some embodiments, the insertion mutation comprises a cysteine insertion mutation. In some embodiments, the insertion mutation comprises a serine insertion mutation. In some embodiments, the insertion mutation comprises a threonine insertion mutation. In some embodiments, the insertion mutation comprises an alanine insertion mutation. In some embodiments, the insertion mutation is located at amino acid position 1079. In some embodiments, the insertion mutation is located at amino acid position 1125. In some embodiments, the insertion mutation is located at amino acid position 1130. In some embodiments, the insertion mutation is located at amino acid position 1133. In some embodiments, the insertion mutation is located at amino acid position 1140. In some embodiments, the insertion mutation is located at amino acid position 1168. In some embodiments, the insertion mutation is located at amino acid position 1173. In some embodiments, the insertion mutation is located at amino acid position 1180. In some embodiments, the insertion mutation is located at amino acid position 1186. In some embodiments, the insertion mutation is located at amino acid position 1203. In some embodiments, the insertion mutation is located at amino acid position 1212. In some embodiments, the insertion mutation is located immediately before D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212 of the *S. Pyogenes* Cas9.

[0130] The Cas nickase may comprise a sequence of any one of SEQ ID NO: 85 – SEQ ID NO: 87 or SEQ ID NO: 90 – SEQ ID NO: 92. In some embodiments, the Cas nickase comprises a sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to any one of SEQ ID NO: 85 – SEQ ID NO: 87 or SEQ ID NO: 90 – SEQ ID NO: 92. The Cas nickase may be expressed as two or more segments. The cysteine point mutation may be located at the N-terminus of a C-

terminal portion of the Cas nickase. A first segment may comprise a sequence of SEQ ID NO: 90.

[0131] Disclosed herein are compositions comprising a Cas nickase and a reverse transcriptase that do not include heterodimerization domains. For example, the Cas nickase and reverse transcriptase may not be engineered to heterodimerize. And yet, the Cas nickase and reverse transcriptase may heterodimerize and perform nucleic acid editing without said engineering or heterodimerization domains. The Cas nickase and the reverse transcriptase may include separate polypeptide chains.

[0132] Disclosed herein are compositions comprising a Cas nickase and a reverse transcriptase, or at least one polynucleotide encoding the Cas nickase and reverse transcriptase, wherein at least part of the Cas nickase and the reverse transcriptase are included in at least 2 separate polypeptide chains, wherein the at least 2 separate polypeptide chains comprise separate polypeptide chains comprising heterodimer domains that bind one another. The separate polypeptide chains may include fusion proteins comprising the heterodimer domains. The heterodimer domains may be fused to the separate polypeptide chains. The heterodimer domains may be fused to amino or carboxy ends of the separate polypeptide chains. A heterodimer domain may include a leucine zipper. A heterodimer domain may include a PDZ domain. A heterodimer domain may include streptavidin. A heterodimer domain may include a streptavidin binding protein. A heterodimer domain may include a foldon domain. A heterodimer domain may include a hydrophobic polypeptide. A heterodimer domain may include an antibody. A heterodimer domain may include a knob, a hole, a leucine zipper, a coiled coil, or a polar amino acid residue capable of forming an electrostatic interaction. A heterodimer domain may include any of heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), immunoglobulin Fc region, heavy chain domain 3 (CH3) of IgG or IgA, heavy chain domain 4 (CH4) of IgM or IgE, Fab, Fab₂, leucine zipper motifs, barnase-barstar dimers, miniantibodies, or ZIP miniantibodies. A heterodimer domain may include a fibrin foldon domain. A heterodimer domain may include a leucine zipper, foldon domain, fragment X, collagen domain, 2G12 IgG homodimer, mitochondrial antiviral-signaling protein CARD filament, Cardiac phospholamban transmembrane pentamer, parathyroid hormone dimerization domain, Glycophorin A transmembrane, HIV Gp41 trimerisation domain, or HPV45 oncoprotein E7 C-terminal dimer domain. A heterodimer domain may include an Fc domain. A heterodimer domain may include a leucine zipper domain, PSD95-Dlg1-zo-1 (PDZ) domain, streptavidin, streptavidin binding protein (SBP), FKBP binding domain (FRB) of mTOR, Cyclophilin-Fas fusion protein (CyP-Fas), Calcineurin A (CNA) and FK506 binding protein (FKBP), Snap tag, Halo tag, PYL or

ABI. A heterodimer domain may include a binding fragment of a heterodimer domain described herein.

Expression of Split Cas9 Constructs using Intein Technology

[0133] In some embodiments, a Cas9 construct (e.g., a Cas9-RT) may be expressed as split construct as one or more exteins fused to one or more inteins. Intein technology may be used to deliver large proteins into a cell by expressing the protein as two or more shorter peptide segments (exteins). Each extein may be expressed as a fusion with an intein peptide (e.g., an Npu C intein or an Npu N intein). An intein may autocatalyze fusion of two or more exteins and may autocatalyze excision of the intein from its corresponding extein. The result may be a protein complex comprising a first extein fused to a second extein and lacking inteins. An intein may be positioned N-terminal of the extein, or an intein may be positioned C-terminal of the extein. An extein may comprise a cysteine residue positioned adjacent to the intein (e.g., at the C-terminal end of an extein with an intein fused to the C-terminal end of the extein). The Cas nickase may be expressed as two or more segments. A first of the Cas nickase segment may comprise an N-terminal portion of the Cas nickase. A first segment of the Cas nickase may comprise a first intein. A second segment of the Cas nickase may comprise a C-terminal portion of the Cas nickase. A second segment of the Cas nickase may comprise a second intein. An intein may be fused to a C-terminus of an N-terminal portion of the Cas nickase. An intein may be fused to an N-terminus of a C-terminal portion of the Cas nickase.

[0134] A nucleic acid sequence encoding an extein-intein fusion may fit into a delivery vector (e.g., an adeno-associated virus (AAV) vector). In some embodiments, a vector encoding a peptide segment extein fused to an intein may be delivered to a cell. In some embodiments, the extein-intein fusion may be expressed in a cell. A first extein-intein fusion peptide may be fused to one or more additional extein-intein fusion peptide, and the inteins may be excised to produce a large protein construct lacking inteins. In some embodiments, a protein may comprise a point mutation to introduce a cysteine residue to facilitate extein fusion and intein excision. In some embodiments, a Cas9-RT of the present disclosure may be expressed as two or more extein-intein fusion peptides. In some embodiments, a Cas9 of the present disclosure (e.g., SEQ ID NO: 32 – SEQ ID NO: 37 or SEQ ID NO: 84 – SEQ ID NO: 87) may be expressed in conjunction with a reverse transcriptase of the present disclosure (e.g., SEQ ID NO: 3 – SEQ ID NO: 22 or SEQ ID NO: 40 – SEQ ID NO: 80) as two or more extein-intein fusion peptides to produce a Cas9-RT fusion. For example, a Cas9-RT may be expressed as a first Cas9-RT extein-fusion comprising nCas9(1-1172) – Npu N intein and a second Cas9-RT extein-fusion comprising Npu C intein – nCas9(1173-1368) – mlvRT5M. nCas9(1-1172) may correspond to residues 1-1172 of

a nicking Cas9 (e.g., residues 1-1172 of any one of SEQ ID NO: 84 – SEQ ID NO: 87). nCas9(1173-1368) may correspond to residues 1173-1368 of a nicking Cas9 with a cysteine at position 1173 (e.g., residues 1-1172 of SEQ ID NO: 86). mlvRT5M may correspond to a reverse transcriptase comprising 5 point mutations (e.g., SEQ ID NO: 13 or SEQ ID NO: 40). A segment may comprise a sequence of SEQ ID NO: 91. The segment may comprise a reverse transcriptase fused to a Cas nickase (e.g. the C-terminal portion of the Cas nickase). The reverse transcriptase may comprise an N-terminus fused to a C-terminus of the C-terminal portion of the Cas nickase. The reverse transcriptase may comprise an mlvRT, or a variant thereof.

Guide Nucleic Acids

[0135] Provided herein are guide nucleic acids (e.g., gRNAs) that direct a programmable nuclease (e.g., a nCas9) to a target nucleic acid. A guide nucleic acid of the present disclosure may facilitate synthesis of a nucleic acid sequence to be inserted into a target site of the target nucleic acid. A guide nucleic acid of the present disclosure may facilitate editing of a nucleic acid sequence at a target site of the target nucleic acid.

[0136] In some embodiments, a guide nucleic acid of the present disclosure may comprise a spacer reverse complementary to a first region of a target nucleic acid, a scaffold configured to bind to a Cas nickase, a reverse transcriptase template encoding a sequence to be incorporated into the target nucleic acid (RTT), a first strand primer binding site reverse complementary to a second region of the target nucleic acid, a second strand primer comprising a sequence of a region of the reverse transcriptase template, or a combination thereof. In some embodiments, the first region of the target nucleic acid is on a first strand of the target nucleic acid and the second region of the target nucleic acid is on the second strand of the target nucleic acid. In some embodiments, all or part of the first region of the target nucleic acid is reverse complementary to all or part of the second region of the target nucleic acid. In some embodiments, the first strand primer binding site is configured to hybridize to the second region of the target nucleic acid. In some embodiments, the reverse transcriptase template is configured to serve as a template for reverse transcription from a 3' end of the second region of the target nucleic acid. In some embodiments, the second strand primer is configured to serve as a primer for transcription from a template reverse complementary to the reverse transcriptase template. In some embodiments, the first synthesized strand may be the template for synthesis of a second strand from the second strand primer.

[0137] A guide nucleic acid of the present disclosure may comprise an RTT. This way, a nucleic acid sequence that gets inserted may have a mutation in the PAM. This can prevent re-editing of an already inserted nucleic acid sequence. The RTT may comprise a modification that disrupts a

protospacer adjacent motif (PAM) sequence. The RTT may comprise two or more modifications that disrupt one or more PAM sequences. The modification may comprise a sequence that is partially complementary with the PAM sequence. The modification may comprise a mismatch with the PAM sequence. The PAM sequence may be disrupted in a target nucleic acid. The target nucleic acid may include a naturally occurring PAM sequence prior to the disruption. The PAM sequence may comprise a 2-6 base pair nucleic acid sequence. An example of PAM sequences is 5'-NGG-3'. Other examples of PAM sequences include 5'-TTTN-3' or 5'-YTN-3'. Any of these PAM sequences may be modified in the RTT. Some examples of such modifications may include an insertion, a deletion, or a point mutation. A PAM sequence may be recognized by a Cas nickase. A modified or disrupted PAM sequence may not be recognized by the Cas nickase in some cases. The modification may comprise a sequence that disrupts or eliminates the PAM in the genome.

[0138] The reverse transcriptase template may comprise a modification that disrupts a mononucleotide track in the genome. The modification may comprise a sequence that is partially complementary with the mononucleotide track. The modification may comprise a mismatch with the mononucleotide track. The reverse transcriptase template may comprise two or more modifications that disrupt one or more mononucleotide tracks in the genome. The modification may comprise a sequence that disrupts or eliminates the mononucleotide track in the genome. The guide nucleic acid may comprise one or more modifications in the reverse transcriptase template that eliminate one or more tracks of at least 4 consecutive nucleotides of the same base in the target nucleic acid.

[0139] A target nucleic acid may include polyA tracks or long polyA tracks. An RTT may include long of polyA tracks. Introducing a modification in the RTT to disrupt the polyA track may improve an editing efficiency. In some embodiments, the RTT further comprises one or more modifications that eliminate or modify tracks of at least 4 consecutive nucleotides that are the same nucleotide base. The one or more modifications in the reverse transcriptase template may eliminate one or more tracks of consecutive nucleotides (e.g. at least 4 consecutive nucleotides) of the same base in the target nucleic acid. The RTT may comprises a modification that eliminates 4 or more consecutive A nucleotides. The RTT may comprises a modification that eliminates 4 or more consecutive T nucleotides. The RTT may comprises a modification that eliminates 4 or more consecutive G nucleotides. The RTT may comprises a modification that eliminates 4 or more consecutive C nucleotides. The RTT may comprises a modification that eliminates 4 or more consecutive U nucleotides. The RTT may comprises a modification that eliminates 3 or more consecutive nucleotides, wherein the 3 or more consecutive nucleotides all

comprise the same nucleobase as each other. The RTT may comprises a modification that eliminates 4 or more consecutive nucleotides, wherein the 4 or more consecutive nucleotides all comprise the same nucleobase as each other. The RTT may comprises a modification that eliminates 5 or more consecutive nucleotides, wherein the 5 or more consecutive nucleotides all comprise the same nucleobase as each other. The RTT may comprises a modification that eliminates 6 or more consecutive nucleotides, wherein the 6 or more consecutive nucleotides all comprise the same nucleobase as each other. The RTT may comprises a modification that eliminates 7 or more consecutive nucleotides, wherein the 7 or more consecutive nucleotides all comprise the same nucleobase as each other. The RTT may comprises a modification that eliminates 8 or more consecutive nucleotides, wherein the 8 or more consecutive nucleotides all comprise the same nucleobase as each other. The RTT may comprises a modification that eliminates 5, 6, 7, 8, 9, 10, or more consecutive nucleotides that include the same base. The RTT may comprises a modification that modifies 4, 5, 6, 7, 8, 9, 10, or more consecutive nucleotides that include the same base in a row to no longer comprise the consecutive nucleotides that include the same base in a row.

[0140] The modification may comprise a mutation in relation to an unmodified guide nucleic acid. The mutation may be a silent mutation. In some cases, the mutation is not a silent mutation.

[0141] The guide nucleic acid may comprise a region that binds to itself another region on the guide nucleic acid to improve gene editing. A guide nucleic acid of the present disclosure may comprise a Velcro region. The Velcro region may comprise a region of the guide nucleic acid that binds to another region of the guide nucleic acid referred to as a “Velcro binding site.” For example, the Velcro region may comprise a region of the guide nucleic acid that binds to another region of the guide nucleic acid to improve gene editing. The binding of the Velcro region to a Velcro binding site may alter a structure of the guide nucleic acid. The altered structure of the guide nucleic acid by the binding of the Velcro region to the Velcro binding site may improve gene editing. The guide nucleic acid may comprise a gRNA positioning system (GPS). The Velcro region or GPS may hybridize to a region of the gRNA (e.g. a Velcro binding site or a GPS binding site). The Velcro region or GPS may hybridize to a region of the reverse transcriptase template. “Velcro” and “GPS” may be used interchangeably. For example, a “Velcro region” may be referred to as a “GPS region,” or vice versa; or a “Velcro binding site” may be referred to as a “GPS binding site,” or vice versa. The Velcro region may hybridize to a region of the reverse transcriptase template region. A gRNA comprising a Velcro region may include a second strand primer. A gRNA comprising a Velcro region may comprise a spacer, a scaffold region, a Velcro region, a RT template, a SSP, a ribozyme, or a combination thereof.

For example, a gRNA comprising a Velcro region may comprise a spacer, a scaffold region, a Velcro region, a RT template, a SSP, and a ribozyme. A gRNA comprising a Velcro region may comprise a spacer, a scaffold region, a RT template, a Velcro region, a SSP, a ribozyme, or a combination thereof. For example, a gRNA comprising a Velcro region may comprise a spacer, a scaffold region, a RT template, a Velcro region, a SSP, and a ribozyme. A gRNA comprising a Velcro region may comprise a spacer, a scaffold region, a RT template, a Velcro region, a SSP, a ribozyme, or a primer binding site (PBS), or a combination thereof. For example, a gRNA comprising a Velcro region may comprise a spacer, a scaffold region, a RT template, a Velcro region, a SSP, a ribozyme, and a PBS. Examples of gRNAs comprising Velcro regions are shown in **FIG. 8A** and **FIG. 8B** and in **FIG. 12A** and **FIG. 12B**. The Velcro region may facilitate reverse transcription of a nucleic acid sequence to be inserted into a target nucleic acid at a target site.

[0142] The guide nucleic acid comprising a GPS region may comprise a guide RNA. The guide nucleic acid comprising a GPS region may comprise a guide nucleic acid other than a guide RNA. The guide nucleic acid comprising a GPS binding site may comprise a guide RNA. The guide nucleic acid comprising a GPS binding site may comprise a guide nucleic acid other than a guide RNA.

[0143] The Velcro region may be synthetic. A Velcro binding site may be synthetic. The Velcro region and the Velcro binding site may be inserted into a gRNA. For example, a synthetic Velcro binding site may be included in a gRNA 5' of a RT template in a gRNA.

[0144] Disclosed herein, in some embodiments, are Velcro regions or Velcro binding sites. In some embodiments, a nucleic acid comprises a Velcro region. In some embodiments, one or more viral vectors (e.g. adenoviruses) comprises the nucleic acid comprising the Velcro region. In some embodiments, a cell comprises the nucleic acid comprising the Velcro region. In some embodiments, a guide nucleic acid comprises a Velcro region. The Velcro region may hybridize to a Velcro binding site. In some embodiments, the Velcro binding site is reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is at least 50% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is at least 60% reverse complementary to the Velcro region. In some

embodiments, the Velcro binding site is at least 70% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is at least 80% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is at least 90% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is 100% reverse complementary to the Velcro region. In some embodiments, the reverse transcriptase template region comprises the Velcro binding site. In some embodiments, the Velcro binding site is 3' of a primer binding site (e.g. a first strand primer binding site, or a second strand primer binding site). In some embodiments, the Velcro binding site is 5' of a primer binding site. In some embodiments, the Velcro region is 3' of a reverse transcriptase template. In some embodiments, the Velcro region is 5' of a reverse transcriptase template. In some embodiments, the Velcro region is 5' of a scaffold. In some embodiments, the Velcro region is 3' of a scaffold. In some embodiments, the scaffold is complementary to a target nucleic acid (e.g. a CFTR nucleic acid, a USH2A nucleic acid, an ABCA4 nucleic acid, an ATP7B nucleic acid, or an HTT nucleic acid). In some embodiments, a synthetic Velcro sequence is inserted between the scaffold and RTT that binds to a sequence that is rev comp to a synthetic Velcro binding site that is inserted after the PBS. In some embodiments, a Velcro region binds to another Velcro region. In some embodiments, the Velcro region hybridizes to a region of a guide nucleic acid that is not the PAM-proximal 20 nucleotides of the spacer sequence.

[0145] In some embodiments, the Velcro binding site is partially reverse complementary to the Velcro region. Perfect complementarity may, in some cases, contribute to truncated AAV genomes, so introducing some bulges or imperfect complementarity may help retain a benefit of GPS without disrupting AAV packaging. AAV genome packaging may in, some instances, be disrupted by secondary structures. GPS may introduce a disruptive secondary structure. Therefore, reducing the degree of complementarity between GPS and the GPS binding site offers a route to eliminate disruption of AAV packaging by GPS. In some embodiments, the Velcro binding site is less than 50%, less than 55%, less than 60%, less than 65%, less than 70%, less than 75%, less than 80%, less than 85%, less than 90%, less than 91%, less than 92%, less than 93%, less than 94%, less than 95%, less than 96%, less than 97%, less than 98%, less than 99%, or less than 100% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is less than 80%, less than 85%, less than 90%, less than 91%, less than 92%, less than 93%, less than 94%, less than 95%, less than 96%, less than 97%, less than 98%, less than 99%, or less than 100% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is less than 70% reverse complementary to the Velcro region. In some embodiments, the Velcro binding site is less than 80% reverse complementary to the Velcro

region. In some embodiments, the Velcro binding site is less than 90% reverse complementary to the Velcro region. Some embodiments include a range of reverse complementarity defined by any two percentages disclosed herein.

[0146] Non-limiting exemplary configurations of a Velcro region are shown in **FIG. 26**. The GPS region may be in any configuration within the guide nucleic acid. The GPS region may be at a 5' end of the guide nucleic acid. The GPS region may be 5' to a spacer. The GPS region may be 5' to and adjacent to a spacer. The GPS region may be 5' to a scaffold. The GPS region may be 5' to and adjacent to a scaffold. The GPS region may be 5' to an RTT. The GPS region may be 5' to and adjacent to an RTT. The GPS region may be 5' to a PBS. The GPS region may be 5' to and adjacent to a PBS. The GPS region may be at a 3' end of the guide nucleic acid. The GPS region may be 3' to a spacer. The GPS region may be 3' to and adjacent to a spacer. The GPS region may be 3' to a scaffold. The GPS region may be 3' to and adjacent to a scaffold. The GPS region may be 3' to an RTT. The GPS region may be 3' to and adjacent to an RTT. The GPS region may be 3' to a PBS. The GPS region may be 3' to and adjacent to a PBS. The GPS region may be within a scaffold. The GPS region may be within an RTT. The GPS region may be within a PBS. Some embodiments include a second GPS region. The second GPS region may be at any of the aforementioned positions. The second GPS region may hybridize to a second GPS binding site. The GPS region may hybridize to the second GPS region.

[0147] A GPS region may comprise a length of nucleotides. For example, the GPS region may be 5-100 nucleotides in length, or about 5-100 nucleotides in length. The GPS region may be 10-50 nucleotides in length, or about 10-50 nucleotides in length. The GPS region may include 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100, or more nucleotides, or a range of nucleotides defined by any two of the aforementioned numbers. The GPS region may include at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 nucleotides. In some cases, the GPS region includes no more than 5, no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 35, no more than 40, no more than 45, no more than 50, no more than 55, no more than 60, no more than 65, no more than 70, no more than 75, no more than 80, no more than 85, no more than 90, no more than 95, or no more than 100 nucleotides. The GPS region may be 20 nucleotides long. The GPS region may be about 20 nucleotides long.

[0148] The GPS region may hybridize to a GPS binding site. The GPS region may be complementary to the GPS binding site. The GPS region may be 100% complementary to the GPS binding site. The GPS region may be at least 50%, at least 60%, at least 70%, at least 80%,

at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, complementary to the GPS binding site. The GPS region may be less than 50%, less than 60%, less than 70%, less than 80%, less than 85%, less than 90%, less than 91%, less than 92%, less than 93%, less than 94%, less than 95%, less than 96%, less than 97%, less than 98%, or less than 99%, complementary to the GPS binding site.

[0149] The GPS region may be at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% complementary to a portion of the GPS binding site. The portion may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 nucleotides. The portion may comprise less than 5, less than 10, less than 15, less than 20, less than 25, less than 30, less than 35, less than 40, less than 45, less than 50, less than 55, less than 60, less than 65, less than 70, less than 75, less than 80, less than 85, less than 90, less than 95, or less than 100 nucleotides.

[0150] The GPS region may be less than 50%, less than 60%, less than 70%, less than 80%, less than 85%, less than 90%, less than 91%, less than 92%, less than 93%, less than 94%, less than 95%, less than 96%, less than 97%, less than 98%, less than 99%, or 100% complementary to a portion (or a second portion) of the GPS binding site. The portion (or second portion) may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 nucleotides. The portion (or second portion) may comprise less than 5, less than 10, less than 15, less than 20, less than 25, less than 30, less than 35, less than 40, less than 45, less than 50, less than 55, less than 60, less than 65, less than 70, less than 75, less than 80, less than 85, less than 90, less than 95, or less than 100 nucleotides.

[0151] In some examples, the GPS region is complementary to 5-10 nucleotides of the GPS binding site. In some examples, the GPS region is at least 80% complementary to 5-10 nucleotides of the GPS binding site. In some examples, the GPS region is complementary to 11-100 nucleotides of the GPS binding site. In some examples, the GPS region is at least 80% complementary to 11-100 nucleotides of the GPS binding site.

[0152] In some embodiments, a gRNA of the present disclosure may comprise a self-cleaving ribozyme, for example as shown in **FIG. 13A**. In some embodiments, a second strand primer region of a gRNA may comprise 100% sequence complementarity to a template region

positioned on the first synthesized strand. In some embodiments, transcription of the second strand primer may produce a poly-U sequence (e.g., UUUUU) at the 3' end of the gRNA and 3' of the second strand primer. Presence of the poly-U sequence immediately 3' of the second strand primer may inhibit function of the second strand primer. A ribozyme sequence may be included in the gRNA to prevent formation of the poly-U sequence immediately 3' of the second strand primer. The ribozyme may autocatalytically cleave itself off of the gRNA. In some embodiments, the ribozyme may be positioned 3' of the second strand primer. The ribozyme positioned 3' of the second strand primer may autocatalytically cleave itself from the gRNA, leaving an in-tact second strand primer without a poly-U sequence. Inclusion of a ribozyme (e.g., an HDV ribozyme) 3' of the second strand primer may enable 100% complementarity of the second strand primer to the template without formation of a poly-U sequence immediately 3' of the second strand primer that inhibits second strand primer function. In some embodiments, a gRNA comprising a self-cleaving ribozyme may have the self-cleaving ribozyme sequence positioned 3' of the second strand primer. In some embodiments, the ribozyme (e.g., an HDV ribozyme) may leave a 2'3' cyclic phosphate at the 3' end of the gRNA following autocatalytic cleavage of the ribozyme. The 2'3' cyclic phosphate may inhibit function of the second strand primer. The 2'3' cyclic phosphate may be converted to a 3' hydroxyl using a polynucleotide kinase. In some embodiments, the polynucleotide kinase is an endogenous polynucleotide kinase present in a cell expressing a gRNA. In some embodiments, the polynucleotide kinase is exogenously expressed.

[0153] In some embodiments, a tRNA may be fused to the gRNA in place of the ribozyme to prevent formation of the poly-U sequence immediately 3' of the second strand primer. In some embodiments, the tRNA may be positioned 3' of the second strand primer. An RNase P enzyme may cleave the tRNA from the rest of the gRNA sequence. In some embodiments, the RNase P may cleave the tRNA from the 3' end of the second strand primer, leaving a 3' hydroxyl at the 3' end of the second strand primer. In some embodiments, the RNase P is an endogenous RNase P present in a cell expressing the gRNA. In some embodiments, the RNase P is exogenously expressed. In some embodiments, a gRNA comprising a tRNA may have the tRNA sequence positioned 3' of the second strand primer. The tRNA may have a sequence corresponding to any tRNA recognized by RNase P. In some embodiments, the tRNA may comprise a sequence of SEQ ID NO: 94.

[0154] The guide nucleic acid may include a spacer. The spacer may be reverse complementary to a first region of a target nucleic acid. The guide nucleic acid may include a scaffold. The scaffold may bind a Cas nickase. The guide nucleic acid may include a reverse transcriptase

template. The reverse transcriptase template may encode a sequence to be inserted into a target nucleic acid. The guide nucleic acid may include a first strand primer binding site. The first strand primer binding site may be reverse complementary to a second region of the target nucleic acid. The guide nucleic acid may comprise a second strand primer. The second strand primer may include a sequence of a region of the reverse transcriptase template.

[0155] Disclosed herein, in some embodiments, are guide nucleic acids comprising a scaffold. The scaffold may bind a nuclease. The scaffold may bind a Cas nuclease. The scaffold may bind a nickase. The scaffold may bind a Cas nickase. The scaffold may bind an *S. Pyogenes* Cas9 nuclease. The scaffold may bind an *S. Pyogenes* Cas9 nickase. The scaffold may include a scaffold nucleic acid sequence. The scaffold nucleic acid sequence may include the sequence of SEQ ID NO: 139. The scaffold nucleic acid sequence may include a sequence that is at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the sequence of SEQ ID NO: 139.

[0156] Disclosed herein, in some embodiments, are guide nucleic acids (e.g. a gRNA). The guide nucleic acid may comprise an extension. The extension may be at a 5' end of the guide nucleic acid. The extension may be at a 3' end of the guide nucleic acid. The guide nucleic acid may comprise a scaffold comprising the extension. The extension may be on a 3' end of the scaffold. The extension may comprise a reverse transcriptase template. The extension may comprise a primer binding site. The extension may contain a reverse transcriptase template and a primer binding site. The primer binding site of the extension may hybridize to a genomic flap generated by a nuclease or nickase. The extension may be oriented using a Velcro region. The extension may be oriented using a Velcro binding site. The extension may be oriented using a Velcro region and a Velcro binding site. The extension may comprise the Velcro region. The Velcro region (and, for example, binding of the Velcro region to the Velcro binding site) may spatially orient the primer binding site to be near the genomic flap. The extension may comprise the Velcro binding region. The guide nucleic acid may include a Velcro region outside of the extension. The guide nucleic acid may include a Velcro binding region outside of the extension.

[0157] The first region of the target nucleic acid may be on a first strand of a target nucleic acid. The second region of the target nucleic acid may be on a second strand of the target nucleic acid. All of the first region of the target nucleic acid may be reverse complementary to all of the second region of the target nucleic acid. All of the first region of the target nucleic acid may be reverse complementary to part of the second region of the target nucleic acid. Part of the first region of the target nucleic acid may be reverse complementary to all of the second region of the

target nucleic acid. Part of the first region of the target nucleic acid may be reverse complementary to part of the second region of the target nucleic acid.

[0158] The guide nucleic acid may comprise a cleavable sequence. The cleavable sequence may be at a 3' end of the guide nucleic acid. The cleavable sequence may be at a 5' end of the guide nucleic acid. The cleavable sequence may comprise a ribozyme cleavable sequence. The cleavable sequence may comprise a tRNA cleavable sequence. The gRNA may include a self cleaving ribozyme such as a HDV ribozyme. The ribozyme may be 3' of a second strand primer (SPP). A tRNA (e.g. a human glutamate tRNA) may be incorporated after the SPP in place of the ribozyme, and this may increase editing efficiency more than with the ribozyme.

[0159] The first strand primer binding site may hybridize to the second region of the target nucleic acid. The reverse transcriptase template may serve as a template for reverse transcription. The reverse transcription may be from a 3' end of the second region of the target nucleic acid. The second strand primer may serve as a primer for transcription from a template. The template may be reverse complementary to the reverse transcriptase template. A first synthesized strand may serve as a template for synthesis of a second strand from the second strand primer.

[0160] Second strand primers (SPPs) may be included in the gRNA. The SPP may be about 10-30 or about 15-25 nucleotides in length. The SPP may be about 20 nucleotides in length. Including a SPP of, for example 20 nucleotides, may increase the efficiency of editing (by 2x, or from about 20% to about 40%, for example). The SPP may be 20, 40, or 60 nucleotides in length, or a range of nucleotides in length defined by any of the aforementioned numbers of nucleotides. Some embodiments include a nucleic acid (e.g. DNA) strand with a desired edit that is complementary to a first strand. This may allow RT to use the first strand as a template. The terminal 3 nucleotides of the SPP may be complementary to the first strand. The SPP may hybridize to a portion of a first strand that is 3' to an edit site. The SPP may be coded to remove secondary structure, and thereby increase editing efficiency.

[0161] A composition described herein may include a first guide nucleic acid. The composition may include a second guide nucleic acid. The second guide nucleic acid may comprise a guide nucleic acid described herein. The first guide nucleic acid may bind to a first Cas nickase. The second guide nucleic acid may bind to a second Cas nickase. A first spacer of the first guide nucleic acid may bind a first Cas nickase. A second spacer of the second guide nucleic acid may bind a second Cas nickase. A first scaffold of the first guide nucleic acid may bind the second Cas nickase. A second scaffold of the second guide nucleic acid may bind the first Cas nickase. The first guide nucleic acid may comprise a first linker. The second guide nucleic acid may comprise a second linker. The first linker may hybridize to the second linker.

[0162] The guide nucleic acid may include gRNA 2.0. The guide nucleic acid may include a 13 nucleotide PBS. The guide nucleic acid may include a 10-15 nucleotide PBS. The guide nucleic acid may include a 13 nucleotide RTT. The guide nucleic acid may include a 10-15 nucleotide RTT. The RTT may encode a mutation as compared to the target nucleic acid.

[0163] Disclosed herein, in some embodiments, are a first and second guide nucleic acid. In some embodiments, the first guide nucleic acid comprises a reverse transcriptase template (RTT). In some embodiments, the second guide nucleic acid comprises a reverse transcriptase template. The reverse transcriptase templates of the first and second guide nucleic acids may be at least partly complementary. In some embodiments, part of the reverse transcriptase template of the second guide nucleic acid is complementary to part of the reverse transcriptase template of the first guide nucleic acid. In some embodiments, the reverse transcriptase template of the second guide nucleic acid is complementary to part of the reverse transcriptase template of the first guide nucleic acid. In some embodiments, part of the reverse transcriptase template of the second guide nucleic acid is complementary to the reverse transcriptase template of the first guide nucleic acid. In some embodiments, the reverse transcriptase template of the second guide nucleic acid is complementary to the reverse transcriptase template of the first guide nucleic acid. In some embodiments, the reverse transcriptase template of the second guide nucleic acid is complementary (or at least partly complementary) to at least part of the reverse transcriptase template of the first guide nucleic acid. The reverse transcriptase templates of the first and second guide nucleic acids may include overlapping dual extended fRNA's (ODEGs) wherein part of the second gRNA's reverse transcriptase template is reverse complementary to part of the first gRNA's reverse transcriptase template. The parts that are complementary may include at least 5 nucleic acids, at least 10 nucleic acids, at least 20 nucleic acids, at least 30 nucleic acids, at least 40 nucleic acids, at least 50 nucleic acids, at least 60 nucleic acids, at least 70 nucleic acids, at least 80 nucleic acids, at least 90 nucleic acids, at least 100 nucleic acids, or more nucleic acids. The parts that are complementary may in some instances include no more than 5 nucleic acids, no more than 10 nucleic acids, no more than 20 nucleic acids, no more than 30 nucleic acids, no more than 40 nucleic acids, no more than 50 nucleic acids, no more than 60 nucleic acids, no more than 70 nucleic acids, no more than 80 nucleic acids, no more than 90 nucleic acids, no more than 100 nucleic acids, or less nucleic acids.

[0164] The guide nucleic acid may comprise gRNA positioning system (GPS). The GPS may include an RNA sequence that binds to a portion of the guide nucleic acid. This may bring the PBS into close proximity with a 5' end of the gRNA. The RNA sequence of the GPS may be 10, 15, 20, 25, or more nucleotides in length, or a range of nucleotides in length defined by any two

of the aforementioned integers. A benefit of using GPS may include increasing editing efficiency when using a long RTT (e.g. an RTT of at least 20, 50, or 100 nucleotides).

[0165] The RNA sequence of the GPS may be about 20 nucleotides in length. The RNA sequence of the GPS may hybridize to a portion of the RTT. The portion of the RTT that the RNA sequence of the GPS hybridizes to may be 10, 15, 20, 25, or more nucleotides in length, or a range of nucleotides in length defined by any two of the aforementioned integers. The portion of the RTT that the RNA sequence of the GPS hybridizes to may be about 20 nucleotides. The portion of the RTT that the RNA sequence of the GPS hybridizes to may be designed to be the same or about the same length as the GPS, or vice versa.

[0166] The GPS may include a version 1 GPS. The guide nucleic acid may include an RNA sequence inserted 5' of the RTT. The RNA sequence may hybridize with the RTT. The RNA sequence may hybridize with a 3' region of the RTT.

[0167] The GPS may include a version 2 GPS. The guide nucleic acid may include an RNA sequence inserted 3' of a PBS. The RNA sequence may hybridize with the RTT. The RNA sequence may hybridize with a 5' portion of the RTT.

[0168] In some cases, the guide nucleic acid comprises one guide nucleic acid, or one type of guide nucleic acid. In some cases, the guide nucleic acid comprises only one guide nucleic acid, or only one type of guide nucleic acid. In some cases, the guide nucleic acid comprises more than one guide nucleic acid, or more than one type of guide nucleic acid. In some cases, the guide nucleic acid comprises two guide nucleic acids, or two types of guide nucleic acid. In some cases, the guide nucleic acid comprises only two guide nucleic acid, or only two types of guide nucleic acid.

[0169] Some aspects of the present disclosure include a single guide nucleic acid system. In some cases, a single guide nucleic acid system might generate a flap containing the desired edit that does not efficiently displace the original genomic strand that doesn't contain the edit. A composition or method for promoting hybridization of the extended flap into the genome may anchor the 3' end of the extended flap in the vicinity of the genomic strand it is intended to replace. GPS-assisted reachover gRNAs (GARGs) may enable this. The GARG may anchor an extended flap. The GARG may anchor a 3' end of an extended flap. In some embodiments, a guide nucleic acid comprises a GARG.

[0170] An example of a GARG is shown in **FIG. 32**. **FIG. 32** shows a GARG that includes a spacer that targets a first region of a target nucleic acid as well as a primer binding site that hybridizes to a second region of the genome that is targeted by a different guide, called a GPS-recruiting guide (GRG). The GARG contains a GPS component that is designed to hybridize to a

GPS-binding site that is part of the GRG. In some embodiments, a guide nucleic acid comprises a GRG. Some embodiments comprise a system including a GARG and GRG. The system may include other gene editing components such as those described herein.

[0171] A guide nucleic acid may comprise a GARG. The GARG may include a spacer. The spacer of the GARG may bind a first region of a target nucleic acid. The spacer of the GARG may be reverse complementary to the first region of a target nucleic acid. The GARG may include an RTT. The RTT may encode a sequence to be inserted into the target nucleic acid. The GARG may include a scaffold. The scaffold may bind to a Cas nuclease, or be configured to bind to a Cas nuclease. The GARG may include a primer binding site (e.g. a first primer binding site). The primer binding site of the GARG may bind to a region of the target nucleic acid that does not include any part of the region of the nucleic acid targeted or bound by the spacer or the nucleic acid reverse complementary to the nucleic acid targeted or bound by the spacer. The primer binding site of the GARG may bind to a region of the target nucleic acid that does not comprise any part of a first region of a target nucleic acid complementary to a spacer of the GARG, and that does not comprise any part of a reverse complement of the first region. The primer binding site of the GARG may bind a second region of the target nucleic acid. The primer binding site of the GARG may be reverse complementary to second region of the target nucleic acid. The GARG may include a GRG binding site. The GRG binding site may bind to a second guide nucleic acid (where the GARG comprises a first guide nucleic acid). The second guide nucleic acid may comprise a GRG. The GRG binding site may bind to a GRG. The GRG binding site may be reverse complementary to a portion of a GRG. The portion of the GRG that is reverse complementary to the GRG binding site may be referred to as a GARG-binding portion. The GARG may comprise a GPS region. The GRG binding site may be the GPS region. The GRG may comprise a GPS binding site. The GARG-binding portion may be the GPS binding site. The second guide nucleic acid may bring the primer binding site into proximity with a genomic flap. The second guide nucleic acid may bring the primer binding site into contact with a genomic flap. The second guide nucleic acid may bring the primer binding site into close proximity with a genomic flap. The inclusion of a GPS region and GPS binding site may pull the end of the GARG to where it may bind the genomic flap. The GARG may be encoded by a nucleic acid such as DNA. Any of the components of the GARG may be included in the GRG. The GRG may be encoded by a nucleic acid such as DNA. The GARG and the GRG be encoded by the same nucleic acid, or by separate nucleic acids. The GARG, or a nucleic acid encoding the GARG, may be encompassed by a virus particle such as an AAV. The GRG, or a nucleic

acid encoding the GRG, may be encompassed by a virus particle such as an AAV. The GARG and the GRG be encompassed by the same virus particle, or by separate virus particles.

[0172] Some embodiments include a dual guide system. The dual guide system may comprise a GARG and a GRG. Some embodiments include a composition comprising a GARG and a GRG. Some embodiments include a method of using a GARG and a GRG, or a method of gene editing with a GARG and a GRG.

[0173] Some embodiments include a gene editing method comprising administering a GARG and a GRG to a cell. Some embodiments include a gene editing method comprising administering one or more nucleic acids that express a GARG and a GRG to a cell. Some embodiments include a gene editing method comprising expressing a GARG and a GRG to a cell. Some embodiments include a gene editing method comprising expressing or administering a GARG to a cell comprising a GRG. Some embodiments include a gene editing method comprising expressing or administering a GRG to a cell comprising a GARG. Some embodiments include a gene editing method comprising expressing a GARG or a GRG in a cell comprising a gene editing enzyme. Some embodiments include a gene editing method comprising expressing a GARG and a GRG in a cell comprising a gene editing enzyme. Some embodiments include a gene editing method comprising administering a GARG or a GRG to a cell comprising a gene editing enzyme. Some embodiments include a gene editing method comprising administering a GARG and a GRG to a cell comprising a gene editing enzyme. The administering may be to a subject comprising the cell.

[0174] Disclosed herein, in some aspects, are compositions or systems comprising an RNA (or polynucleotide) comprising a spacer, a reverse transcriptase template comprising a desired edit, and a primer binding site, in which the primer binding site binds to a nucleic acid that is targeted by a separate RNA. Disclosed herein are systems comprising an RNA or polynucleotide comprising a spacer, a reverse transcriptase template comprising a desired edit, and a primer binding site, in which the primer binding site binds to a nucleic acid that does not comprise any part of the region of the nucleic acid targeted or bound by the spacer or the nucleic acid reverse complementary to the nucleic acid targeted or bound by the spacer.

Compositions for Genome Editing

[0175] Compositions of the present disclosure may facilitate efficient editing of a target nucleic acid at a target site. A composition of the present disclosure may comprise a guide nucleic acid, a nCas9, and a reverse transcriptase. A composition of the present disclosure may comprise a sequence encoding a guide nucleic acid, a nCas9, a reverse transcriptase, or a combination thereof. The nCas9 and the reverse transcriptase may be a fused nCas9-RT construct. The nCas9

and the reverse transcriptase may be a split nCas9-RT construct. A composition of the present disclosure may be introduced into a cell comprising the target nucleic acid, thereby editing the target nucleic acid. In some embodiments, a sequence (e.g., a plasmid) encoding one or more components of the composition may be introduced into a cell comprising the target nucleic acid. The one or more components of the composition may be expressed in the cell. In some embodiments, a composition of the present disclosure may comprise a first guide nucleic acid, a first nCas9s, a first reverse transcriptase, a second guide nucleic acid, a second nCas9s, and a second reverse transcriptase. In some embodiments, the first guide nucleic acid binds to the first nCas9, and the second guide nucleic acid binds to the second nCas9. In some embodiments, a first spacer of the first guide nucleic acid binds the first nCas9, a second spacer of the second guide nucleic acid binds the second nCas9, a first scaffold of the first guide nucleic acid binds the second nCas9, and a second scaffold of the second guide nucleic acid binds the first nCas9. In some embodiments, the first guide nucleic acid comprises a first linker and the second guide nucleic acid comprises a second linker. In some embodiments, the first linker hybridizes to the second linker.

[0176] A composition comprising a first guide nucleic acid and a second guide nucleic acid may facilitate synthesis or editing of a sequence. A composition comprising a first guide nucleic acid and a second guide nucleic acid may facilitate editing of a target nucleic acid at a target site. A composition comprising a first guide nucleic acid and a second guide nucleic acid may be a two single guide system. A composition comprising a first guide nucleic acid and a second guide nucleic acid may be a dual guide system. In a two single guide system, each gRNA binds to a different nCas9 and the two gRNAs each comprise a reverse transcriptase template region. In a dual guide system, each gRNA may bind to a different nCas9. In a two single guide system, only one of the gRNAs may comprise a reverse transcriptase template region. In a two single guide system, the second guide may nick the opposite strand. In a dual guide system, only one of the gRNAs may comprise a reverse transcriptase template region. In a dual guide system, the second guide may nick the opposite strand. In a dual guide complex, the spacer of the first gRNA may bind the first nCas9, the spacer of the second gRNA may bind the second nCas9, the scaffold of the first gRNA may bind the second nCas9, and the scaffold of the second gRNA may bind the first nCas9.

[0177] The guide nucleic acid may form a complex with a Cas nickase. The guide nucleic acid may form a complex with a reverse transcriptase. Upon complex formation, the Cas nickase may introduce a single-strand break at a target site in a target nucleic acid.

[0178] Some non-limiting examples of target nucleic acids include a cystic fibrosis transmembrane conductance regulator (CFTR) nucleic acid, an usherin (USH2A) nucleic acid, an ATP-binding cassette subfamily A member 4 (ABCA4) nucleic acid, a Wilson disease protein (ATP7B) nucleic acid, or a Huntingtin (HTT) nucleic acid. In some embodiments, the target nucleic acid comprises a CFTR gene. In some embodiments, the target nucleic acid comprises a USH2A gene. In some embodiments, the target nucleic acid comprises a ABCA4 gene. In some embodiments, the target nucleic acid comprises a ATP7B gene. In some embodiments, the target nucleic acid comprises a HTT gene.

[0179] Disclosed herein are compositions comprising a Cas nickase, a reverse transcriptase, and a guide nucleic acid. A first polypeptide may comprise the Cas nickase. A second polypeptide may comprise the reverse transcriptase. The guide nucleic acid may bind to the Cas nickase. The guide nucleic acid may bind to the reverse transcriptase.

[0180] The RT may comprise an MS2 coat protein (MCP) peptide. In some cases, the RT does not include an MS2 coat protein (MCP) peptide. For example, the composition may include RWa1. The guide nucleic acid may comprise a MS2 hairpin. In some cases, the guide nucleic acid does not include a MS2 hairpin. The MCP peptide may bind an MS2 hairpin in the guide nucleic acid. The MS2 hairpin may be between a gRNA scaffold and a RTT. This may bring the RT into close proximity with the gRNA to allow editing. A benefit of using a MCP peptide and MS2 hairpin is to separate the RT and Cas nickase (or a portion of them), and allow them to fit within AAV vectors. The MCP peptide and MS2 hairpin may not be necessary. The composition including the MCP peptide or the MS2 hairpin may have an editing efficiency of at least about 3% or 4%, for example, when transfected into cells. The composition including the MCP peptide or the MS2 hairpin may have an editing efficiency of at least about 10% or 15%, for example, when transfected into cells.

[0181] The RT and Cas nickase may include leucine zippers. For example, the composition may include RWb1. The composition including leucine zippers may have an editing efficiency of at least about 35% or 40%, for example, when transfected into cells. The composition including leucine zippers may have an editing efficiency of at least about 3% or 4%, for example, when transduced into cells. A benefit of using leucine zipper is to separate the RT and Cas nickase (or a portion of them), and allow them to fit within AAV vectors. However, the leucine zippers may not be necessary.

[0182] The RT or Cas nickase may be split, for example, using intein splitting. In some cases, the RT and Cas nickase are not split using intein splitting. An example of using intein splitting is RWc1. The split may be between residues 1172 and 1173 of the Cas nickase. The composition

using the split RT or Cas nickase may have an editing efficiency of at least about 25% or 30%, for example, when transfected into cells. A benefit of using RWc1 or a similar splitting method may be to allow for more space for additional nucleotide sequences such as regulatory elements that may be allowed to fit within an AAV vector with a nucleic acid sequence encoding the RT or Cas nickase. For example, the splitting method may allow for about 500, 600, or 700 (or a range defined by any of the aforementioned integers) more nucleotides for additional nucleotide sequences to fit within an AAV.

[0183] The RT or Cas nickase may be separate and not bound together. An example of using non-bound RT and Cas nickase is RWd1.

[0184] A composition of the present disclosure may comprise a protein complex or a sequence encoding a protein complex. The protein complex may comprise a protective protein complex. The protein complex may prevent deamination or degradation of a guide nucleic acid. For example, a protective complex may be a Human Orf1p (SEQ ID NO: 38) or a Murine Orf1p (SEQ ID NO: 39).

[0185] Disclosed herein are methods of increasing genome editing efficiency. The method may include delivering an Orf1p to a cell. The cell may express a composition or a guide nucleic acid described herein.

[0186] Disclosed herein are nucleic acids comprising nucleotide sequences encoding a composition or a guide nucleic acid described herein. Disclosed herein are viral vectors comprising the nucleic acids. Disclosed herein are cells comprising a composition described herein. Disclosed herein are cells comprising a nucleic acid described herein. Disclosed herein are cells comprising a guide nucleic acid described herein. Disclosed herein are cells comprising a viral vector described herein. The cell may be a prokaryotic cell. The cell may be a eukaryotic cell.

[0187] Some embodiments include method of increasing genome editing efficiency by increasing the dNTP concentration such as dNTP concentration in a cell. Inhibiting SAMHD1 may increase the dNTP concentration. Administering dNTPs may increase the dNTP concentration in the cell. Some embodiments include a method of increasing genome editing efficiency comprising inhibiting SAMHD1 in a cell. Some embodiments include a method of increasing genome editing efficiency comprising administering dNTPs to a subject or to a cell

[0188] In some embodiments, a composition of the present disclosure may comprise a protein, a nucleic acid encoding the protein, or a non-coding nucleic acid, for increasing editing efficiency of a Cas9 construct of the present disclosure (e.g., a split Cas9-RT construct). In some embodiments, a protein, a nucleic acid encoding the protein, or a non-coding nucleic acid for

increasing editing efficiency of a Cas9 construct may comprise a protein or a nucleic acid that inhibits the dNTP cleavage activity of SAMHD1 or a nucleic acid encoding a protein that inhibits the dNTP cleavage activity of SAMHD1. For example, a nucleic acid that inhibits the dNTP cleavage activity of SAMHD1 may comprise a microRNA that degrades SAMHD1 transcripts.

[0189] Some embodiments include increasing the dNTP concentration in the cell, relative to a baseline dNTP concentration. In some embodiments, the dNTP concentration is increased by 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or more, or by a range of percentages defined by any two of the aforementioned percentages, relative to the baseline dNTP measurement. In some embodiments, the dNTP concentration is increased by at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, relative to the baseline dNTP measurement. In some embodiments, the dNTP concentration is increased by no more than 5%, no more than 10%, no more than 15%, no more than 20%, no more than 25%, 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%, no more than 100%, relative to the baseline dNTP measurement.

[0190] In various aspects, the low dNTP concentration comprises a dNTP concentration of 0.5 micromolar, 0.6 micromolar, 0.7 micromolar, 0.8 micromolar, 0.9 micromolar, 1.0 micromolar, or 1.1 micromolar, or a range defined by any two of the aforementioned dNTP concentrations. In various aspects, the low dNTP concentration comprises a dNTP concentration of about 0.5 micromolar, about 0.6 micromolar, about 0.7 micromolar, about 0.8 micromolar, about 0.9 micromolar, about 1.0 micromolar, or about 1.1 micromolar, or a range defined by any two of the aforementioned dNTP concentrations. In various aspects, the low dNTP concentration comprises a dNTP concentration below about 1.1 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration below about 1.0 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration below about 0.9 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration below about 0.8 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration below about 0.7 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration below about 0.6 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration below about 0.5 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration above about 0.9 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration above about 0.8 micromolar. In

various aspects, the low dNTP concentration comprises a dNTP concentration above about 0.7 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration above about 0.6 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration above about 0.5 micromolar. In various aspects, the low dNTP concentration comprises a dNTP concentration above about 0.4 micromolar.

[0191] In various aspects, the present disclosure provides a method of increasing gene editing efficiency in a cell having a low deoxynucleoside triphosphate (dNTP) concentration and comprising a DNA polymerase, the method comprising: increasing the dNTP concentration in the cell, relative to a baseline dNTP concentration. In various aspects, increasing the dNTP concentration in the cell comprises inhibiting a deoxynucleotide triphosphate triphosphohydrolase in the cell. In various aspects, the deoxynucleotide triphosphate triphosphohydrolase comprises SAM domain and HD domain-containing protein 1 (SAMHD1). In various aspects, inhibiting SAMHD1 comprises contacting the SAMHD1 with a Vpx protein, or expressing the Vpx protein in the cell. In various aspects, inhibiting SAMHD1 comprises contacting the SAMHD1 with a BGLF4 protein, or expressing the BGLF4 protein in the cell. In various aspects, inhibiting SAMHD1 comprises contacting an mRNA encoding the SAMHD1 with a microRNA or siRNA that hybridizes to the mRNA, or expressing the microRNA or siRNA in the cell. In various aspects, inhibiting SAMHD1 comprises contacting the SAMHD1 with a small molecule SAMHD1 inhibitor. In various aspects, increasing the dNTP concentration in the cell comprises administering dNTPs to the cell. In various aspects, administering dNTPs to the cell comprises administering dNTPs to a subject comprising the cell. In various aspects, increasing the dNTP concentration in the cell comprises administering nucleosides or nucleotides to the cell. The nucleosides or nucleotides may include deoxynucleosides (dNs), deoxynucleoside monophosphates (dNMPs), or nucleoside triphosphates (NTPs). In some cases, the nucleosides or nucleotides are not dNTPs, or do not include dNTPs. In various aspects, administering nucleosides or nucleotides to the cell comprises administering the nucleosides or nucleotides to a subject comprising the cell. In various aspects, the administration is oral or by injection. In various aspects, increasing the dNTP concentration in the cell comprises delivering a dNTP synthetic enzyme to the cell. In various aspects, the dNTP synthetic enzyme comprises a kinase. In various aspects, the kinase comprises a nucleoside kinase, deoxynucleoside kinase, deoxynucleoside monophosphate kinase, or deoxynucleotide diphosphate kinase. In various aspects, the DNA polymerase comprises a reverse transcriptase. The DNA polymerase may be adapted for gene editing. The DNA polymerase may be a gene editing polymerase. The DNA polymerase may be a recombinant DNA polymerase. Some embodiments include introducing the

DNA polymerase into the cell. Some embodiments include expressing the DNA polymerase in the cell. In various aspects, the cell comprises or further comprises a Cas9 programmable nuclease, a guide nucleic acid, or a combination thereof. Some embodiments include introducing into the cell, or expressing the Cas9 programmable nuclease in the cell. Some embodiments include introducing into the cell, or expressing the guide nucleic acid in the cell. The Cas9 programmable nuclease may be part of the DNA polymerase, or may associate with the DNA polymerase. In various aspects, the low dNTP concentration comprises a dNTP concentration found in a nondividing cell. In various aspects, the low dNTP concentration is less than a dNTP concentration found in an activated peripheral blood mononuclear cell. In various aspects, the low dNTP concentration comprises a dNTP concentration below 1 micromolar. In various aspects, the increasing the dNTP concentration comprises increasing the dNTP concentration by at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, or more, relative to the baseline dNTP measurement. In various aspects, the dNTP concentration comprises a deoxyadenosine triphosphate (dATP) concentration, a deoxycytidine triphosphate (dCTP) concentration, a deoxyguanosine triphosphate (dGTP) concentration, or a deoxythymidine triphosphate (dTTP) concentration, or any combination thereof.

[0192] In various aspects, the present disclosure provides a method of increasing gene editing efficiency in a cell having a low deoxynucleoside triphosphate (dNTP) concentration, comprising: increasing the dNTP concentration in the cell, wherein the cell comprises a Cas9 programmable nuclease, a reverse transcriptase, and a guide nucleic acid. In various aspects, the present disclosure provides a method of increasing gene editing efficiency in a cell having a low deoxynucleoside triphosphate (dNTP) concentration, comprising: contacting the cell with a gene editing enzyme modified for efficient catalysis in the low dNTP concentration, or expressing the gene editing enzyme in the cell. In some aspects, increasing the dNTP concentration in the cell comprises inhibiting SAMHD1 in the cell. In some aspects, inhibiting SAMHD1 comprises contacting the SAMHD1 with a Vpx protein, or expressing the Vpx protein in the cell. In some aspects, inhibiting SAMHD1 comprises contacting an mRNA encoding the SAMHD1 with a microRNA or siRNA that hybridizes to the mRNA, or expressing the microRNA or siRNA in the cell. In some aspects, inhibiting SAMHD1 comprises contacting the SAMHD1 with a small molecule SAMHD1 inhibitor. In some aspects, increasing the dNTP concentration in the cell comprises administering dNTPs to the cell. In some aspects, increasing the dNTP concentration in the cell comprises delivering a dNTP synthetic enzyme to the cell. In some aspects, the dNTP synthetic enzyme comprises a deoxynucleoside diphosphate (dNDP) kinase. In some aspects, the

gene editing enzyme comprises a Cas9 programmable nuclease or a reverse transcriptase. In some aspects, the reverse transcriptase is modified by introducing a point mutation at position Q84, L139, Q221, V223, T664, or L671. In some embodiments, the method further comprises measuring the dNTP concentration. Some embodiments include measuring a dNTP concentration after increasing the dNTP concentration, and determining an increase relative to a baseline dNTP concentration.

[0193] Some embodiments include obtaining determining the increase in dNTP concentration. Some embodiments include measuring the dNTP concentration. In some embodiments, the dNTP concentration is measured using an assay such as an absorbance assay, a colorimetric assay, or an enzyme-linked immunosorbent assay. Some embodiments include measuring the baseline dNTP concentration. In some embodiments, the baseline dNTP concentration is measured using an assay such as an absorbance assay, a colorimetric assay, or an enzyme-linked immunosorbent assay.

[0194] Disclosed herein are methods of increasing genome editing efficiency. The method may include inhibiting SAMHD1 in a cell. The cell may express a Cas9 programmable nuclease. The cell may express a Cas nickase. The cell may express a reverse transcriptase. The cell may express a guide nucleic acid. An example of inhibiting SAMHD1 may include treating the cell with a SAMHD1 inhibitor such as a small molecule SAMHD1 inhibitor. An example of inhibiting SAMHD1 may include expressing a microRNA against SAMHD1 in the cell.

[0195] A protein for increasing editing efficiency may be a Vpx protein (e.g., SEQ ID NO: 82, SEQ ID NO: 83, or SEQ ID NO: 93). Vpx is in some instances a lentiviral protein. Vpx is in some instances a immunodeficiency virus (SIV) protein which may be used for increasing editing efficiency (e.g. by inhibiting SAMHD1). A Vpx protein may increase editing efficiency of a Cas9-RT construct by increasing the availability of dNTPs in a cell. For example, a Vpx protein may inhibit the dNTP cleavage activity of SAMHD1, thereby increasing availability of dNTPs in the cell. In some embodiments, a Vpx protein may be co-expressed in a cell with a Cas9-RT construct of the present disclosure. The Cas9-RT construct expressed with the Vpx protein may have increased editing efficiency compared to the Cas9-RT construct in the absence of the Vpx protein. A Vpx peptide may be a Hiv2-rod Vpx (e.g., SEQ ID NO: 82). In some embodiments, a Vpx protein may be expressed as its own coding sequence. In some embodiments, a Vpx protein may be expressed in the same coding sequence as the reverse transcriptase. For example, a Vpx protein may be expressed in the same coding sequence as the reverse transcriptase, separated by a p2a self-cleaving peptide (e.g., SEQ ID NON: 83). In some embodiments, a Vpx protein may be a Vpx RH-2-1 D8 protein (e.g., SEQ ID NO: 93). In some

embodiments, a Vpx protein may be expressed in the same coding sequence as the Cas9 protein. Inhibiting SAMHD1 may comprise expressing a Vpx protein in the cell.

[0196] Disclosed herein are methods of increasing genome editing efficiency comprising expressing a Vpx protein in a cell. The cell may express a composition described herein. The cell may express a guide nucleic acid described herein.

[0197] Some embodiments include a method of increasing a dNTP concentration in a cell, or of improving gene editing, by inhibiting a deoxynucleotide triphosphate triphosphohydrolase (dNTPase) such as SAMHD1. Some embodiments relate to a composition for inhibiting SAMHD1. A Vpx protein may be used to inhibit SAMHD1. A BGLF4 protein may be used to inhibit SAMHD1. BGLF4 may phosphorylate SAMHD1 and thereby inhibit a dNTPase activity of SAMHD1. BGLF4 is in some instances an Epstein-Barr virus (EBV)-encoded protein kinase. An EBV-encoded protein kinase may be used for increasing editing efficiency (e.g. by inhibiting SAMHD1). The composition for inhibiting SAMHD1 may include a small molecule SAMHD1 inhibitor. The small molecule SAMHD1 inhibitor may comprise pppCH₂dU, or a salt thereof. The small molecule SAMHD1 inhibitor may comprise dGMPNPP, or a salt thereof.

[0198] Disclosed herein are methods of increasing genome editing efficiency by increasing the concentration of nucleosides or nucleotides (e.g. dNTPs) in a cell. The cell may express a Cas9 programmable nuclease. The cell may express a Cas nickase. The cell may express a reverse transcriptase. The cell may express a guide nucleic acid. An example of increasing the concentration of dNTPs in a cell comprises delivering nucleotides or nucleosides to a cell. Increasing the concentration of nucleosides or nucleotides in a cell may include delivery of the nucleosides or nucleotides to the cell. The nucleotides or nucleosides may then be converted into dNTPs in the cell. Delivery of the nucleosides or nucleotides may include oral delivery or injection. Conversion of the nucleosides or nucleotides to dNTPs may be through phosphorylation by endogenous kinases or synathesis, for example through endogenous salvage pathways. The method may comprise delivering nucleotides or nucleosides to the cell, resulting in an increased concentration of dNTPs in the cell compared to a cell that did not received the nucleotides or nucleosides. The increased concentration of the dNTPs in the cell may result in increased editing efficiency in the cell comprising the compositions as disclosed herein.

[0199] Disclosed here are methods that include using SAMHD1 overexpression to screen for RT mutants that operate better in limiting dNTP concentrations. Also disclosed are methods for screening or identifying improved RTs in cells that are modified to overexpress SAMHD1 or a unphosphorylatable mutant of SAMHD1. Some embodiments include overexpressing SAMHD1 in cells. Some embodiments include expressing a mutant SAMHD1 that has been mutated to

prevent phosphorylation of a residue of the mutant SAMHD1 in cells. Some embodiments include identifying an RT activity in the cells. Some embodiments include identifying the RT as an improved RT based on the RT activity. Some embodiments include a method for screening or identifying an improved reverse transcriptase (RT), comprising: overexpressing SAMHD1, or expressing a mutant SAMHD1 that has been mutated to prevent phosphorylation of a residue of the mutant SAMHD1, in cells; identifying an RT activity in the cells; and based on the RT activity, identifying the RT as an improved RT.

AAV and Methods for Delivery of Precision Editing Components

[0200] Described herein are precision editing components such as Cas nickases, reverse transcriptases (RTs), or guide RNAs (gRNAs). The nickase and RT may be encoded by polynucleotides. The polynucleotides may be delivered by AAVs. The polynucleotides encoding the nickase and RT may be engineered to fit within the AAVs. Examples are provided herein for engineering the nickase and RT to fit within AAVs. For example, the nickase and RT may be engineered to dimerize. The nickase and RT may be coexpressed. The nickase may be split using an intein system. Part of the nickase may be combined as a fusion protein with the RT. A goal of the exemplary dimerization, coexpression and split intein systems is to be able to deliver the genome editing components using AAVs comprising 4.5 kb carrying capacities.

[0201] **FIG. 24** shows that when plasmids expressing nCas9 and mcp-mlvRT5m were cotransfected in HEK293T-BFP cells with gRNAs that did and did not include an ms2 hairpin, the same BFP to GFP editing efficiency was achieved. As such, coexpression of unfused and non-dimerizing nCas9 and a reverse transcriptase in the same cell can result in editing. Therefore NLS-nCas9 (SEQ ID NO: 138) and mlvRT5m-NLS (SEQ ID NO: 95) may be coexpressed from separate AAVs to achieve efficient editing. Therefore, coexpression of an unfused and non-dimerizing Cas nickase and a reverse transcriptase in the same cell can result in editing. Therefore a Cas nickase and an RT may be coexpressed from separate AAVs to achieve efficient editing. Likewise, a Cas nickase and an RT that have been engineered to dimerize may be coexpressed from separate AAVs to achieve efficient editing.

[0202] A Cas nickase and a RT may be encoded by polynucleotides. The Cas nickase and RT may be encoded by 2 separate polynucleotides, or part of one may be included in the other polynucleotide (for example, as described herein). One or more AAVs may comprising the polynucleotides. At least part of the Cas nickase and RT may be encompassed or comprised within separate AAVs. Part of the Cas nickase and RT may be encompassed or comprised within separate AAVs. All of the Cas nickase and RT may be encompassed or comprised within separate AAVs.

[0203] In some cases, a composition is included, which includes a Cas nickase and a reverse transcriptase, wherein at least part of the Cas nickase and the reverse transcriptase are included in separate polypeptide chains, and wherein the Cas nickase and the reverse transcriptase form a Cas-reverse transcriptase heterodimer. The separate polypeptide chains may be encoded by separate polynucleotides. The separate polynucleotides may be included in separate viral vectors such as AAVs. The separate polynucleotides may be divided into 2, 3, 4, 5, 6, 7, 8, 9, or 10, of the separate viral vectors. The separate polynucleotides may be divided into 2 of the separate viral vectors. The separate polynucleotides may be divided into at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10, of the separate viral vectors. The separate polynucleotides may be divided into no more than 2, no more than 3, no more than 4, no more than 5, no more than 6, no more than 7, no more than 8, no more than 9, or no more than 10, of the separate viral vectors. The separate polynucleotides may be divided into no more than 2 of the separate viral vectors. The separate polynucleotides may be divided into no more than 3 of the separate viral vectors. The separate polynucleotides may be divided into no more than 4 of the separate viral vectors.

[0204] The separate polynucleotides may be short enough to fit within separate AAV genomes (e.g. each below about 4500 bp). For example, separate polynucleotides may each be about the sizes described in **FIG. 19A**. Separate polynucleotides may each be less than or no greater than about the sizes described in **FIG. 19A**. Separate polynucleotides may each be less than or no greater than about 10% less than or greater than the sizes described in **FIG. 19A**. Separate polynucleotides may each be less than about 4500 bp. Separate polynucleotides may include a range of polynucleotide sizes, such as ranges including any of the sizes in **FIG. 19A**, or ranges including about the sizes in **FIG. 19A**.

[0205] In some cases, the AAVs include a first AAV. The first AAV may include a first polynucleotide, which may encode a Cas or Cas component such as a Cas nickase described herein. The AAVs may include a second AAV, which may include a second polynucleotide encoding a RT such as a RT described herein.

[0206] Examples of AAVs may include AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV-DJ, AAV-DJ/8, AAV-Rh10, AAV-Rh74, AAV-retro, AAV-PHP.B, AAV8-PHP.eB, or AAV-PHP.S, or a combination of thereof. Examples of AAVs may include a serotype such as AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12. Examples of AAVs may include a pseudotype such as AAV-DJ, AAV-DJ/8, AAV-Rh10, AAV-Rh74, AAV-retro, AAV-PHP.B, AAV8-PHP.eB, or AAV-PHP.S.

[0207] The AAV may comprise an AAV genome. The AAV genome may comprise pCMV-NLS-nSpCas9(1-1172)-NpuN-cMycNLS-48pA, or any combination of components thereof. An AAV genome comprising pCMV-NLS-nSpCas9(1-1172)-NpuN-cMycNLS-48pA may include the sequence of SEQ ID NO: 142. The AAV genome may include a sequence at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the sequence of SEQ ID NO: 142.

[0208] The AAV genome may comprise pCMV-NpuC-nSpCas9(1173-1368;S1173C)-mlvRT14M-SV40 NLS-P2A-VPXrh21-48pA-pU6-ush2a-gRNA, or any combination of components thereof. An AAV genome comprising pCMV-NpuC-nSpCas9(1173-1368;S1173C)-mlvRT14M-SV40 NLS-P2A-VPXrh21-48pA-pU6-ush2a-gRNA may include the sequence of SEQ ID NO: 143. The AAV genome may include a sequence at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the sequence of SEQ ID NO: 143. The AAV may be useful for treatment of Usher syndrome, or restoring a USH2A mutant. The AAV may be administered to a subject with Usher syndrome.

[0209] Following administration of the AAVs to a subject, genome editing in the subject may be measured or assessed. Following administration of the AAVs to one or more cells, genome editing in the one or more cells may be measured or assessed. Genome editing may be measured or assessed by sequencing. Genome editing may be measured or assessed by an assay. The assay may comprise measuring or identifying an edited genome in a subject or cell. The assay may comprise measuring or identifying an RNA resultant from the edited genome in a subject or cell. The assay may comprise measuring or identifying a protein resultant from the edited genome in a subject or cell. Some examples of an assay include a hybridization assay, an immunoassay, a colorimetric assay, a fluorescent assay, or mass spectrometry.

[0210] Some embodiments include a method for introducing one or more changes in the nucleotide sequence of a DNA molecule at a target locus, comprising: contacting the DNA molecule with a programmable nuclease and a guide nucleic acid which targets the programmable nuclease to the target locus. The programmable nuclease may form a complex with a reverse transcriptase.

Methods of treatment using Precision Editing Components

[0211] Some embodiments include treatment of a genetic disorder using a method or composition described herein. For example, some embodiments include administering one or more nucleic acids comprising or encoding gene editing components described herein. For example, a viral vector may be used to deliver the administered nucleic acids.

[0212] The administration may include injection. The administration may include administration of a composition comprising the nucleic acids. The administration may include administration of a composition comprising the viral vector. The composition may comprise a pharmaceutical composition. The composition may comprise a pharmaceutical composition. The pharmaceutical composition may include a carrier such as water, a buffer, or a saline solution. The pharmaceutical composition may include liposomes.

[0213] The administration may be to a subject in need thereof. For example, the administration may be to a subject having a genetic disorder. The subject may be a vertebrate. The subject may be a mammal. The subject may be a human. In some embodiments, the administration corrects a disease-causing gene mutation in the subject. In some embodiments, the administration corrects a disease-causing gene mutation in a cell of the subject.

[0214] Some non-limiting examples of genetic disorders include adenosine deaminase deficiency, alpha-1 antitrypsin deficiency, cystic fibrosis, a muscular dystrophy (e.g. Duchenne muscular dystrophy), galactosemia, hemochromatosis, Huntington's disease, maple syrup urine disease, Marfan syndrome, neurofibromatosis (e.g. Type 1), pachyonychia congenita, phenylketonuria, severe combined immunodeficiency, sickle cell disease, Smith-Lemli-Opitz syndrome, or Tay-Sachs disease. In some embodiments, the genetic disorder comprises cystic fibrosis, Stargardt disease, Usher syndrome, or Huntington's disease. In some embodiments, the genetic disorder comprises cystic fibrosis. In some embodiments, the genetic disorder comprises Stargardt disease. In some embodiments, the genetic disorder comprises Usher syndrome. In some embodiments, the genetic disorder comprises Huntington's disease. In some embodiments, the genetic disorder comprises a polygenic disorder such as heart disease, high blood pressure, Alzheimer's disease, arthritis, diabetes, cancer, or obesity.

AAV-Deliverable Precision Editing Without Double-Stranded Breaks

Summary

[0215] The ability to precisely edit genomes may have profound implications on healthcare, agriculture, or biological sciences, and precise genome editing may cure genetic diseases. While CRISPR nucleases have democratized the ability to target double stranded breaks, generating precise sequence alterations has been difficult due to the inefficiency of homology-directed repair (HDR) at the site of a toxic double-stranded break (DSB) using foreign homologous DNA. Prime Editors have enabled versatile precision editing without relying on HDR but may utilize components that are too large to be delivered with the gene delivery vehicle adeno-associated virus (AAV), have a limited editing window length, break both strands to achieve efficient editing, or have limited efficiency in non-dividing cells. Here these limitations have been

overcome with a set of tools called Rewriter. Rewriter's split systems may provide four modular architectures to deliver the gene editing components within two AAV genomes. Rewriter's optimized reverse transcriptase and guide RNA positioning system (GPS) may increase the editing efficiency and editing window length while only generating one single-stranded break. Finally, Rewriter's anti-restriction factor promotes editing in non-dividing cells. Rewriter achieved 75% editing within a 65 nucleotide window, the highest efficiency reported to date for targeted multi-nucleotide changes in mammalian cells without generating DSBs. Finally, Rewriter components were developed that precisely edited a genomic site commonly mutated in patients with inherited deafness and blindness with no detectable off-target mutations.

[0216] Described herein are compositions and methods for versatile, efficient, and precise genome editing without homology-directed repair or double-stranded breaks using Rewriter, a dual AAV-deliverable system that utilizes an engineered Cas9 nickase to target an optimized reverse transcriptase for synthesis of defined double stranded DNA contiguous with the genome using a guide RNA positioning system (GPS), a second strand primer, and an anti-restriction factor that provides for efficient editing in nondividing cells. Rewriter installed complex sequence changes with up to 54% efficiency within a 65 nucleotide window. A benefit of this system is that it may be used to correct mutations that cause genetic disorders such as cystic fibrosis or Usher syndrome. Correction of genetic defects in Usher syndrome may be used to prevent or treat deafness or blindness. The safety, efficiency, precision, and versatility of Rewriter may be used to treat diseases, improve foods, or advance basic biologic research.

Introduction

[0217] CRISPR nucleases may offer a straightforward approach to create targeted, double-stranded breaks in genomic DNA. However, precisely altering the sequence of a genomic target has been difficult due to inefficiency of homology-directed repair, toxicity of double-stranded breaks, and the challenge of delivering homologous donor DNA. DSBs can lead to long and imprecise deletions that extend beyond the target gene and can even result in the removal of an entire chromosome. Additionally, the vectors encoding the genome editors themselves can be unintentionally integrated at the site of DSBs. Lastly, even a single DSB can activate the p53 pathway, which can lead to apoptosis. CRISPR-guided nucleotide deaminases, or base editors, can avoid double-stranded breaks, and may not rely on homology-directed repair, but may also be limited to making a subset of all substitutions and can cause genomic and transcriptomic off-target mutations. Prime editing may be a more precise and versatile approach for installing insertions, deletions, and complex sequence changes within a 30-nucleotide (nt) window at the protospacer adjacent motif (PAM)-proximal side of a Cas9 cleavage site. Prime Editors (PEs)

may include a nicking *Streptococcus pyogenes* Cas9 (nSpCas9) that generates a genomic flap that can hybridize to a primer binding site (PBS) in a 3' extension of the CRISPR guide RNA (gRNA). A Moloney leukemia virus reverse transcriptase comprising five point mutations (mlvRT5M), which is fused to the nSpCas9, may utilize the genomic flap as a primer to synthesize a strand of DNA comprising the desired edit according to the sequence of the RT template (RTT) which is also included in the gRNA's 3' extension. Editing efficiency can be increased by expressing a second gRNA that nicks the other strand. Unfortunately, PEs have generally been too large to be delivered with adeno-associated virus (AAV). Additionally, PEs may suffer from limited efficiency in non-dividing cells, generate unintended insertions or deletions when nicking both strands, or have a short editing window length that may limit the number of pathogenic mutations that can be corrected with a single construct and may require a PAM to be available near the edit site. Described herein is a gene editing system, which in some instances is called Rewriter, that may install any type of mutation within a larger window and with a higher efficiency than prime editors in dividing and non-dividing cells without generating double-stranded breaks. This may be done using components delivered with AAV.

Precision editors that fit in AAV

[0218] While prime editor and a split intein prime editor may install many types of mutations at a target site, their coding sequences alone may be too large to fit into a 4.5 kilobase (kb) carrying capacity of an AAV (**FIG. 19A**). To provide benefits of precision editing with components that can be delivered within as few as two AAV genomes, Rewriter a1 ("RWa1" or "RW1M") was developed. RWa1 may include nCas9 coexpressed with a MS2 coat protein (MCP) peptide fused to mlvRT5M and a gRNA comprising an MS2 hairpin to which the MCP peptide may specifically bind. RW1M may incorporate an MS2 hairpin into a gRNA bound by nSpCas9 to recruit an MS2-binding peptide (MBP) fused to mlvRT5M. RW1M may allow delivery of nSpCas9 in one AAV, and MBP-mlvRT5M and a gRNA in another AAV (**FIG. 27**).

[0219] As shown in **FIG. 27**, an editing system comprising PE2 may include a nicking Cas9 (nCas9) fused to a Moloney leukemia virus reverse transcriptase containing 5 point mutations (mlvRT5M). The guide RNA used in PE2 may include an extension on the 3' end of the scaffold sequence containing a reverse transcriptase template (RTT) sequence and a primer binding site (PBS) sequence. The nCas9 may first nick the non-target strand which releases a genomic flap that hybridizes with the PBS. The mlvRT5M may then extend the genomic flap by reverse transcribing the RTT. An editing system comprising split PE2 may utilize an Npu split intein to express two ORFs that catalytically splice together to form the nCas9-mlvRT5M fusion protein. An editing system comprising Rewriter 1M (RW1M) may utilize nCas9, a MS2-binding peptide

(MBP) fused to mlvRT5M and a gRNA containing the MS2 hairpin to which the MBP specifically binds (**FIG. 27**: Left: the MS2 hairpin may be inserted within the gRNA scaffold; Middle: the MS2 hairpin may be inserted between the gRNA scaffold and the RTT; Right: the MS2 hairpin may be inserted after the PBS). An editing system comprising RW1L may utilize heterodimerizing leucine zippers to colocalize nCas9 and mlvRT5M. An editing system comprising RW1I may utilize a novel mutant nCas9 that can be split with Npu inteins to produce a nCas9-mlvRT5M protein using ORFs that each fit in AAV. An editing system comprising RW1N may coexpress nCas9 and mlvRT5M without any engineered recruitment components.

[0220] Precision editing efficiencies were determined using a HEK293 cell line stably expressing BFP, which can be edited to GFP by installing a specific 3-nucleotide (nt) mutation. A plasmid expressing the first protein component (nSpCas9), a second plasmid expressing the second protein component (MBP-mlvRT5M), and a third plasmid expressing a gRNA were transiently cotransfected into the HEK293 cell line. Cotransfection of plasmids expressing RWa1 (nCas9 and MCP-mlvRT5M) with gRNA 2.0, which included MS2 hairpin insertions within the Cas9-binding scaffold, further modified with a 13-nt PBS and a 13-nt RTT encoding a +2 ATGG to CATA mutation that was intended to remove the PAM site and install the GFP mutation resulted in 3.7% GFP+ cells, compared to 19% with PE2 (**FIG. 11B**). Inserting the MS2 hairpin closer to the site of the genomic flap and PBS may promote initiation of reverse transcription by the recruited MBP-mlvRT5M. The editing efficiency of RWa1 was increased to 15.5% by inserting the MS2 hairpin between an unmodified gRNA scaffold and the RTT.

[0221] Rewriter b1 (“RWb1” or “RW1L”) may be an alternative approach to deliver editing components within two AAV genomes. The editing components may include a nCas9 fused to a leucine zipper (nSpCas9-LZ1) that heterodimerizes with another leucine zipper fused to an N-terminus of mlvRT5M (LZ2-mlvRT5M) (**FIG. 19A**). Cotransfection of plasmids expressing RWb1 (nSpCas9-LZ1 and LZ2-mlvRT5M) and gRNA 2.0 comprising a 13-nt PBS and a 13-nt RTT encoding a +2 ATGG to CATA mutation resulted in 38% GFP+ cells (**FIG. 11B**). As shown in **FIG. 11B**, HEK293 cells transfected with PE2 and gRNA 2.0, which may contain MS2 hairpin insertions within the Cas9-binding scaffold, further modified with a 13-nt PBS and a 13-nt RTT encoding a +2 ATGG to CATA mutation intended to remove the PAM site and install the GFP mutation, resulted in an editing efficiency of 19%. RWa1 with the same gRNA resulted in 3.7% GFP+ cells. The editing efficiency of RWa1 was increased to 15.5% by inserting the MS2 hairpin between an unmodified gRNA scaffold and the RTT. RWb1 and the gRNA 2.0 construct comprising a 13-nt PBS and a 13-nt RTT encoding a +2 ATGG to CATA mutation resulted in 38% GFP+ cells. RW1L was developed as an alternative design to deliver the editing

components within two AAV genomes. RW1L may comprise a nSpCas9 fused to a leucine zipper17 (nSpCas9-LZ1) that heterodimerizes with a complementary leucine zipper fused to the N-terminus of mlvRT5M (LZ2-mlvRT5M). To allow direct comparison to RW1M, RW1L was initially tested with the same gRNA containing the gRNA 2.0 scaffold that was constructed for RW1M, (though the MS2 hairpins in gRNA 2.0 were necessarily not expected to interact with the components of RW1L).

[0222] While RWa1 and RWb1 may each include two polypeptides that are within the size constraints permissive to AAV packaging, the size of the nSpCas9 and nSpCas9-LZ1 open reading frames (ORFs) may limit the length of regulatory elements that can be included to control expression. In some instances, due to a possible requirement for a cysteine as a first residue of a Npu C-terminal extein, Split Prime Editor 2 may include inteins to split nSpCas9-mlvRT5M into two fragments at Cys574, the most C-terminal cysteine in nSpCas9. It was considered that if an appropriate intein-flanking sequence could be introduced between residues 700 and 1250 of nSpCas9, then a modified nSpCas9-mlvRT5M fusion protein with two intein-comprising ORFs could be encoded within two recombinant AAV genomes with more room for regulatory elements than RWa1 and RWb1 may otherwise allow. A panel of constructs was tested, which encoded N-terminal fragments of nSpCas9 fused to the Npu N-terminal intein paired with an Npu C-terminal intein fused to a C-terminal fragment of nSpCas9 comprising a cysteine substitution (providing, in some instances, intein catalysis) and mlvRT5M. Splitting a Ser1173Cys nSpCas9-mlvRT5M mutant between residues 1172 and 1173 (nCas9(1-1172)-NpuN and nCas9(1173-1368; S1173C)-mlvRT5M; named “RWc1” or “RW1I”) resulted in about a 2-fold greater editing efficiency than PE2 (29% vs 15%, **FIG. 19B**).

[0223] Using the standard gRNA scaffold, RWa1, RWb1, and RWc1 were all found to result in above 40% editing efficiency (**FIG. 19C**). Surprisingly, coexpression of a gRNA not containing the MS2 hairpins with the nCas9 and MCP-mlvRT5M constructs resulted in approximately the same editing efficiency compared to gRNAs including an MS2 hairpin, showing that recruiting or fusing the RT to the site of the nCas9 may not always be necessary, or that simply coexpressing an nCas9 and RT can result in efficient editing. Constructs that did not actively recruit the mlvRT were called Rewriter d1 (“RWd1” or “RW1N”). In the context of AAV delivery, RWc1 may be used, as it may accommodate up to approximately 670 nt of regulatory elements. Transducing the HEK293 cells expressing BFP with RWc1 packaged into two separate AAV2 constructs at an MOI of 2.8×10^5 for each virus resulted in 74.8% GFP+ cells (**FIG. 19D**).

[0224] To demonstrate AAV-mediated delivery of Rewriter, the N-terminal protein component of RW1I, which accommodated a 584-nt CMV promoter, was packaged into an AAV2 vector while packaging both the C-terminal protein component of RW1I (also driven by a CMV promoter) and a gRNA that converts BFP to GFP driven by a U6 promoter, into a separate AAV2 vector. Simultaneous AAV co-transduction of BFP-expressing HEK293 cells at an MOI of 2.8×10^5 VG/cell for each virus resulted in 74.8% GFP+ cells.

[0225] Results indicated that RW1M, RW1L, RW1I, or RW1N may have different editing efficiencies between different gRNAs, so some embodiments include screening several architectures to maximize efficiency with a given gRNA. RW1I may provide the most room for regulatory elements on both AAV genomes, while RW1L and RW1N may accommodate multiple gRNA cassettes in the second AAV genome. Although RW1N surprisingly provided similar editing efficiency as the other architectures when using plasmid transfection, this might require a high enough intracellular concentration of protein components to obviate a possible need for active recruitment. Therefore, RW1N may not provide high editing efficiencies with alternate delivery strategies and expression levels in some cases.

RNA extension reorientation

[0226] Spatial reorientation of a guide RNA extension may increase editing efficiency or window length. Increasing an editing window length may allow screening more gRNAs for efficient editing of a given mutation, or may correct more pathogenic mutations with a single gRNA. Additionally, Prime Editor 3b may increase editing efficiency by expressing a second gRNA that binds a sequence generated by the RT and nicks the opposite strand, thereby evading nick-mediated mismatch repair of the RT-synthesized edit. The Prime Editor 3b approach may, in some instances, be limited to target sites that have a PAM site within the limited editing window length of prime editors. Increasing the editing window length may require increasing a length of the RTT. It was hypothesized that editing efficiencies of Rewriter and PEs may be limited by a rate of hybridization of the PBS and the genomic flap when using longer RTTs. It was also predicted that if the length of the RTT could be increased, a decrease could be affected in the frequency of reverse transcription of a portion of the gRNA scaffold, which may inadvertently lead to undesirable insertion of a scaffold-templated sequence into the genome. An RNA component was designed, here called the gRNA positioning system (GPS), that may be introduced into the 3' extension of the gRNA to spatially orient the PBS to be near the genomic flap regardless of the length of the RTT. GPS Version 1 (V1) may include an RNA sequence inserted 5' of the RTT that may hybridize with a 3' region of the RTT, and/or GPS Version 2 (V2) may be an RNA sequence inserted 3' of the PBS that may hybridizes with the 5' portion of

the RTT (**FIG. 20A**). Computational RNA folding analysis predicted that GPS V1 and V2 would bring the PBS closer to the 5' end of the gRNA's 3' extension as intended (**FIG. 12B**). RNA folding analysis predicted that GPS V1 and V2 may alter the structure of the gRNA's 3' extension to bring the PBS closer to the gRNA scaffold.

[0227] The GPS may also be referred to as Velcro. Likewise, Velcro may be referred to as GPS. GPS may include Velcro or a component of Velcro. Velcro may include GPS or a component of GPS.

[0228] Cotransfection of RWb1 and a guide RNA comprising a 107-nt RTT resulted in 14% GFP+ cells, which was significantly lower than the 38% achieved using a shorter 13-nt RTT, and thus confirmed that increasing the RTT length to broaden the editing window can decrease editing efficiency. Adding a 20-nt GPS V2 increased the editing efficiency to 27.4%. GPS V1 increased editing efficiency from 14% to 24.8%. GPS V1 may be used to cause reverse transcription and genomic insertion of the GPS sequence. Use of a 20-nt GPS V2, 3' of the PBS, increased the editing efficiency to 27.4% and does not have the potential for genomic insertion of the GPS sequence. GPS V2 was proceeded with, and the Rewriter systems using the GPS V2 component may be referred to as "Rewriter 2.0," or may be denoted with "g" (for example: "RW1I_g"). Next, it was found that installing a 3-nt mutation 65-nt from the site of the nick using a 129-nt template was increased 4-fold by incorporating GPS V2 (**FIG. 20B**). The efficiency of making this edit was increased to 21% by recoding the RTT to remove secondary structure that might inhibit reverse transcription while maintaining the amino acid sequence of the target site. Finally, a panel of GPSs of various lengths and binding sites was generated, and a 20-nt GPS that hybridized to the first 20-nt of the RTT resulted in the highest editing efficiency among the set (**FIG. 12D**). A 20-nt GPS that hybridized to the first 20-nt of the RTT resulted in the highest editing efficiency among a panel of GPS V2s of varying lengths and binding sites. Mutation rate data are shown as mean $\pm$ one standard deviation from three biologically independent samples.

Second strand synthesis

[0229] Next, it was hypothesized that editing efficiency could be further increased by synthesizing a second strand of DNA comprising a desired edit that is complementary to the first synthesized strand. Second strand primers (SSP) were introduced. An SSP may allow the reverse transcriptase to use the first synthesized strand as a template for second strand synthesis (**FIG. 13A**). SSP may be inserted into the 3' terminus of the 3' gRNA extension. SSP may hybridize to a portion of the first synthesized strand that is 3' of the edit site. After the PBS hybridizes to the flap and the first strand is reverse transcribed, the SSP may hybridize to the first synthesized

strand, allowing the reverse transcriptase to use the first synthesized strand as a template for second strand synthesis.

[0230] The SSP may hybridize to a portion of the first synthesized strand that is 3' of the edit site. To allow the region from the start of the SSP to the 3' end of the gRNA to be complementary to the first synthesized strand, a self-cleaving hepatitis delta virus (HDV) ribozyme was introduced 3' of the SSP. First tested was whether SSP could improve the editing efficiency of PE2. 20-nt SSPs were found to perform better than 40-nt or 60-nt SSPs (**FIG. 13B**). 20-nt SSPs performed better than 40- and 60-nt SSPs. Incorporating a 20-nt SSP that hybridized up to 6- and 36-nt from the nick site approximately doubled editing efficiency compared to no SSP. A 20-nt SSP that hybridized up to 55-nt from the nick site did not improve editing efficiency. It was predicted that this was due to a lower efficiency of reverse transcription of the more distal portion of the RTT, thereby limiting the availability of the SSP binding site. The systems that utilize this SSP technology were named Rewriter 3.0. It was then found that SSP could be inserted after Velcro using Rewriter 2.0 to further increase editing efficiency, resulting in ~41% editing (Rewriter 3.2; **FIG. 21**). It was also demonstrated that the increase in efficiency that SSP provided was abolished when the terminal 3-nt of SSP were not complementary to the first synthesized strand.

[0231] Since the HDV ribozyme may leave a 2'3' cyclic phosphate on the 3' terminus, and reverse transcription may use a 3' hydroxyl to initiate synthesis from a primer, an endogenous enzyme such as human polynucleotide kinase may convert the 2'3' cyclic phosphate to a 3' hydroxyl. It was predicted that incorporating a tRNA after the SSP in place of the HDV ribozyme could lead to a more rapid generation of the 3' hydroxyl following RNase P cleavage of the tRNA. Incorporating a human glutamate tRNA led to a statistically significant increase in editing efficiency to 50.9%. The efficiency of making an edit 65-nt from the nick was slightly increased by recoding the RTT to remove secondary structure that might inhibit reverse transcription while maintaining the amino acid sequence of the target site.

Engineering mlvRT

[0232] PE2 was developed by introducing mutations into the mlvRT of PE1 that were reported to improve reverse transcriptase activity *in vitro*. The mlvRT used in PE2 may be improved further by incorporating mutations that increase processivity, thermostability, substrate affinity, or modulate RNaseH activity. Therefore, 31 mutations in mlvRT were screened that may improve mlvRT activity *in vitro* (**FIG. 14A**). Five mutations had statistically significant increases in editing. Effects of combinations of these mutations on top of mlvRT5M were tested to determine potential increases in editing efficiency. By adding 9 mutations to mlvRT5M,

editing efficiency was increased from ~43% for mlvRT5M to ~54% for mlvRT14M (**FIG. 14B**). Systems that incorporate mlvRT14M may be referred to as “Rewriter 4” or as “Rewriter 2” (for example: “RW2I_g”).

Overcoming low dNTP concentrations

[0233] Genome editors that polymerize DNA using reverse transcriptases, such as PE or some embodiments of Rewriter, may use dNTPs as substrates. It is therefore conceivable that low dNTP concentrations characteristic of non-dividing or slowly dividing cells (such as in the retina and lung) could possibly pose a barrier compared to editing in rapidly dividing cells in culture. SAMHD1 is a triphosphohydrolase that may control cellular dNTP concentrations. In nondividing cells, SAMHD1 may hydrolyze dNTPs. In cycling cells, cyclin dependent kinase 1 (CDK1) may phosphorylate SAMHD1. This may inhibit dNTP hydrolysis. This may lead to a higher dNTP concentration. A SAMHD1 T592A mutant (“SAMHD1^P” or “SAMHD1 (T592A)”) may in some instances not be phosphorylated, and therefore may deplete dNTP pools regardless of the presence of CDK1. VPX may be a small protein expressed by HIV-2 to specifically target SAMHD1 for degradation. It was predicted that the lower dNTP concentrations of nondividing and slowly dividing cells in HEK293T cells could be modeled by expressing SAMHD1^P (**FIG. 15A**).

[0234] **FIG. 15B** shows that cotransfecting Rewriter 3.2 (which may also be referred to as “RW1L_g”) with SAMHD1^P decreased editing efficiency 2.7-fold, supporting the idea that dNTP concentrations may in some instances be limiting for genome editing. Editing efficiency decreased 2.7-fold when SAMHD1 (T592A) was coexpressed. Several mutations to mlvRT5M restored some of the editing efficiency in the presence of SAMHD1 (T592A). Some mutations may lower the K_m of mlvRT for dNTPs. Introducing some such mutations into Rewriter restored some of the editing efficiency in the presence of SAMHD1^P 2-nt away from the nick. It was confirmed that one of these mutations, V223A, did not reduce efficiency. The construct that incorporated V223A was named Rewriter 5.0 (**FIG. 15C**). Introducing the V223A mutation into mlvRT5M did not reduce the efficiency of installing an edit 65-nt from the nick (Rewriter 5).

[0235] As a complementary approach to increasing editing efficiency in nondividing and slowly dividing cells, VPX was employed. VPX may be a small HIV-2 protein. VPX may specifically target SAMHD1 for degradation. Coexpression of VPX^{ROD} (from the ROD HIV-2 isolate) and SAMHD1^P completely reversed the reduction in editing efficiency caused by expressing SAMHD1^P without VPX (**FIG. 15D**). The SAMHD1^P-induced decrease in editing efficiency was even more drastic when a mutation was installed 65-nt from the nick. Near zero efficiency may be due to the compounding reduction in DNA synthesis efficiency for every dNTP

incorporation. Coexpression of VPX^{ROD} restored the editing efficiency to 78% of the efficiency observed without expressing SAMHD1^P (Rewriter 5.1).

[0236] A variant of VPX was identified from HIV-2 clinical isolate RH2-1 (VPX^{RH2-1}) that fully restored the efficiency of installing a mutation 65-nt from the nick in the presence of SAMHD1^P (Rewriter 5.2; **FIG. 15E**). Mutation rate data in **FIG. 15E** are shown as mean $\pm$ one standard deviation from three biologically independent samples. VPX^{RH2-1} also outperformed VPX^{ROD} when coexpressing SAMHD1 (T592A) with RW2I and a gRNA that may install a mutation 2-nt from the nick, yielding 64% editing efficiency (**FIG. 15F**), as well as with RW2I_g and a gRNA installing a mutation 65-nt from the nick (**FIG. 23**). Rewriter systems that incorporate VPX^{RH2-1} may be designated with “_v” (for example: “RW2I_gv”).

[0237] Increasing concentration of dNTPs may be an additional complementary approach to increase editing efficiency in a cell. The method may comprise delivering nucleotides or nucleosides to the cell, resulting in an increased concentration of dNTPs in the cell compared to a cell that did not received the nucleotides or nucleosides. The increased concentration of the dNTPs in the cell may result in increased editing efficiency in the cell comprising the compositions as disclosed herein. In some cases, dNTPs are administered to a subject (e.g. a subject comprising the cell). In some embodiments, administering dNTPs to a cell comprises administering the dNTPs to a subject comprising the cell. The administration of dNTPs may include oral administration. The administration of dNTPs may be by injection.

Correcting cystic fibrosis mutations or other disease mutations

[0238] Small-molecule therapies for cystic fibrosis have shown a less than complete functional restoration of sweat chloride, pulmonary complication rate, and forced expiratory volume. Additionally, a majority of cystic fibrosis (CF)-causing mutations in the CFTR gene may not be treatable by these small molecules. Delivering a functional copy of a CFTR gene to affected cells may treat any CF patient, regardless of the CF patient's CFTR genotype. However, gene therapy approaches for treating CF have been limited by transient and synthetic regulation of CFTR expression, as well as the limited packaging capacity of AAV, which may require use of a truncated CFTR that displays incomplete activity. Some data related to *CFTR* editing are shown in **FIG. 29B** and **FIG. 29C**.

[0239] As an alternative, editing CF-causing mutations to restore CFTR activity that is controlled by its natural regulatory elements may provide long term and potentially curative therapy. As such, some embodiments of the methods and compositions described herein may be used to treat CF. Some embodiments include administering to a subject in need thereof (e.g. a subject with CF), one or more polynucleotides encoding genome editing components described

herein that are configured to correct a mutant CFTR gene, or one or more viruses such as adenoviruses comprising the one or more polynucleotides. An example of such a component includes a guide RNA comprising a spacer that is reverse complementary to a region of a CFTR nucleic acid. The spacer include the nucleic acid sequence of SEQ ID NO: 96. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 96. The spacer include the nucleic acid sequence of SEQ ID NO: 97. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 97. The spacer include the nucleic acid sequence of SEQ ID NO: 98. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 98. The spacer include the nucleic acid sequence of SEQ ID NO: 99. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 99. The spacer include the nucleic acid sequence of SEQ ID NO: 100. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 100. The spacer include the nucleic acid sequence of SEQ ID NO: 101. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 101. The spacer include the nucleic acid sequence of SEQ ID NO: 102. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 102. The spacer include the nucleic acid sequence of SEQ ID NO: 103. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 103. The spacer include the nucleic acid sequence of SEQ ID NO: 104. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 104. The spacer include the nucleic acid sequence of SEQ ID NO: 105. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 105. In some embodiments, the administration improves a therapeutic parameter of CF in the subject.

TABLE 1 – Spacers

Spacer name (gene.mutation.number)	Sequence	Editing efficiency
CFTR.F508del.1	SEQ ID NO: 96	6.81%
CFTR.F508del.2	SEQ ID NO: 97	
CFTR.F508del.3	SEQ ID NO: 98	
CFTR.R553X.1	SEQ ID NO: 99	13.30%

CFTR.G542X.NAG.1	SEQ ID NO: 100	
CFTR.G542X.NAG.2	SEQ ID NO: 101	
CFTR.G542X.NAG.3	SEQ ID NO: 102	
CFTR.W1282X.1	SEQ ID NO: 103	18.47%
CFTR.W1282X.2	SEQ ID NO: 104	
CFTR.W1282X.3	SEQ ID NO: 105	
USH2A.1	SEQ ID NO: 106	41%
USH2A.2	SEQ ID NO: 107	
ABCA4.G1961E.1	SEQ ID NO: 108	37.2%
ABCA4.G1961E.2	SEQ ID NO: 109	23.3%
ABCA4.G863A.1	SEQ ID NO: 110	
ABCA4.G863A.2	SEQ ID NO: 111	22.45%
ATP7B.H1069Q.1	SEQ ID NO: 112	26.7%
ATP7B.H1069Q.2	SEQ ID NO: 113	10.37%
ATP7B.R778L.1	SEQ ID NO: 114	
ATP7B.R778L.2	SEQ ID NO: 115	
HTT.NAG.1	SEQ ID NO: 116	25%
HTT.2	SEQ ID NO: 117	
HTT.3	SEQ ID NO: 118	
HTT.4	SEQ ID NO: 119	

[0240] Likewise, some embodiments of the methods and compositions described herein may be used to treat other diseases. Editing a disease-causing mutation to restore a correct sequence in a target gene may provide long term and potentially curative therapy for subjects with the disease. Some embodiments include administering to a subject in need thereof (e.g. a subject with a disease), one or more polynucleotides encoding genome editing components described herein that are configured to correct a mutant target gene, or one or more viruses such as adenoviruses comprising the one or more polynucleotides. An example of such a component includes a guide RNA comprising a spacer that is reverse complementary to a region of a target nucleic acid. Some such spacers are included in **TABLE 1**. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or 100%, identical to a spacer in **TABLE 1**, or a complementary sequence. In some cases, the spacer includes a similar sequence but includes Us in place of Ts. In some embodiments, the administration improves a therapeutic parameter of the disease in the subject.

[0241] Some embodiments of the methods and compositions described herein may be used to treat Usher syndrome. Editing Usher syndrome-causing mutations to restore USH2A may provide long term and potentially curative therapy for subjects with Usher syndrome. Some embodiments include administering to a subject in need thereof (e.g. a subject with Usher syndrome), one or more polynucleotides encoding genome editing components described herein

that are configured to correct a mutant *USH2A* gene, or one or more viruses such as adenoviruses comprising the one or more polynucleotides. An example of such a component includes a guide RNA comprising a spacer that is reverse complementary to a region of a *USH2A* nucleic acid. The spacer include the nucleic acid sequence of SEQ ID NO: 106. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 106. The spacer include the nucleic acid sequence of SEQ ID NO: 107. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 107. In some embodiments, the administration improves a therapeutic parameter of Usher syndrome in the subject. In some embodiments, the genome editing components correct a mutation shown in **FIG. 25B**. Some data related to *USH2A* editing are shown in **FIG. 29A**.

[0242] The *USH2A*.1 spacer was used to generate the data in **FIG. 25A** and **FIG. 25B**. For the data in **FIG. 25A**, HEK293T cells were transfected with RW2I using various sequences of PBS, RTT, or GPS. Editing efficiency is displayed as the percentage of reads with the intended 2298T>C mutation. A 15-nt PBS, 52-nt RTT, and 20-nt GPS resulted in 22% editing. Similar experiments were also performed where editing efficiencies are included in **TABLE 1**. For the data in **FIG. 25B**, the most frequent mutant allele generated by the 15-nt PBS, 52-nt RTT, 20-nt GPS construct contained both of the 2298T>C and 2316C>A mutations encoded in the RTT (18.8%). An additional 7.1% of reads were represented by either the 2316C>A PAM-disrupting mutation alone or the target 2298T>C mutation alone. A low frequency of adenine insertions were also detected along poly-A tracts within the target sequence. Data shown include mean $\pm$ one standard deviation from three biologically independent samples.

[0243] Some embodiments of the methods and compositions described herein may be used to treat Stargardt disease. Editing Stargardt disease-causing mutations to restore *ABCA4* may provide long term and potentially curative therapy for subjects with Stargardt disease. Some embodiments include administering to a subject in need thereof (e.g. a subject with Stargardt disease), one or more polynucleotides encoding genome editing components described herein that are configured to correct a mutant *ABCA4* gene, or one or more viruses such as adenoviruses comprising the one or more polynucleotides. An example of such a component includes a guide RNA comprising a spacer that is reverse complementary to a region of an *ABCA4* nucleic acid. The spacer include the nucleic acid sequence of SEQ ID NO: 108. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 108. The spacer include the nucleic acid sequence of SEQ ID NO: 109. The spacer may include a nucleic acid sequence at least 70%, at

least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 109. The spacer include the nucleic acid sequence of SEQ ID NO: 110. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 110. The spacer include the nucleic acid sequence of SEQ ID NO: 111. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 111. In some embodiments, the administration improves a therapeutic parameter of Stargardt disease in the subject. Some data related to *ABCA4* editing are shown in **FIG. 29D-29F**.

[0244] Some embodiments of the methods and compositions described herein may be used to treat Wilson disease. Editing Wilson disease-causing mutations to restore ATP7B may provide long term and potentially curative therapy for subjects with Wilson disease. Some embodiments include administering to a subject in need thereof (e.g. a subject with Wilson disease), one or more polynucleotides encoding genome editing components described herein that are configured to correct a mutant ATP7B gene, or one or more viruses such as adenoviruses comprising the one or more polynucleotides. An example of such a component includes a guide RNA comprising a spacer that is reverse complementary to a region of an ATP7B nucleic acid. The spacer include the nucleic acid sequence of SEQ ID NO: 112. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 112. The spacer include the nucleic acid sequence of SEQ ID NO: 113. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 113. The spacer include the nucleic acid sequence of SEQ ID NO: 114. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 114. The spacer include the nucleic acid sequence of SEQ ID NO: 115. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 115. In some embodiments, the administration improves a therapeutic parameter of Wilson disease in the subject.

[0245] Some embodiments of the methods and compositions described herein may be used to treat Huntington's disease. Editing Huntington's disease-causing mutations to restore HTT may provide long term and potentially curative therapy for subjects with Huntington's disease. Some embodiments include administering to a subject in need thereof (e.g. a subject with Huntington's disease), one or more polynucleotides encoding genome editing components described herein that are configured to correct a mutant HTT gene, or one or more viruses such as adenoviruses comprising the one or more polynucleotides. An example of such a component includes a guide

RNA comprising a spacer that is reverse complementary to a region of an HTT nucleic acid. The spacer include the nucleic acid sequence of SEQ ID NO: 116. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 116. The spacer include the nucleic acid sequence of SEQ ID NO: 117. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 117. The spacer include the nucleic acid sequence of SEQ ID NO: 118. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 118. The spacer include the nucleic acid sequence of SEQ ID NO: 119. The spacer may include a nucleic acid sequence at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 119. In some embodiments, the administration improves a therapeutic parameter of Huntington's disease in the subject.

[0246] Disclosed herein are guide nucleic acids including an extension. The extension may include a extension nucleic acid sequence for editing HTT. The extension nucleic acid sequence for editing HTT may include the sequence of SEQ ID NO: 140. The extension nucleic acid sequence for editing HTT may include a sequence that is at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the sequence of SEQ ID NO: 140. The extension may include a GPS region. The extension may not include a GPS region. An example of an extension sequence of an extension including a GPS region is included in SEQ ID NO: 141. The extension nucleic acid sequence for editing HTT may include the sequence of SEQ ID NO: 141. The extension nucleic acid sequence for editing HTT may include a sequence that is at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the sequence of SEQ ID NO: 141.

Rewriting an Usher syndrome gene

[0247] Usher syndrome may be the most common inherited source of combined deafness and vision loss. While hearing aids and ear implants can treat deafness in some Usher patients, their vision loss is currently untreatable. Possibly the most common mutation that causes Usher syndrome is a single-nucleotide deletion, 2299delG, in the *USH2A* gene. Rewriter may offer an approach for providing curative treatment for patients with Usher syndrome, as gene therapy involving delivery of functional *USH2A* may otherwise hampered by a *USH2A* cDNA size (15.6 kb) beyond the typical AAV and lentivirus packaging capacity, and base editors may otherwise be unable to perform nucleotide insertions.

[0248] Components were provided to rewrite the 2299G region in *USH2A*. Editing efficiency was determined in wildtype HEK293T cells by quantifying the installment of a 2298T>C silent

mutation that was encoded in the RTT 23-nt from the nick. Although 2299G is encoded by the RTT, 2298T>C was also included as a surrogate mutation because wildtype HEK293T cells already contain 2299G. Also encoded was a 2316C>A PAM-disabling silent mutation in the RTT 5-nt from the nick that was intended to prevent nSpCas9 from continuing to nick the target site after the intended edit is achieved.

[0249] Twenty-two percent editing was achieved through the use of a 15-nt PBS, 52-nt RTT, and 20-nt GPS in a RW2I system (**FIG. 25A**). For the data in **FIG. 25A**, HEK293T cells were transfected with RW2I using different sequences of PBS, RTT, and GPS. Editing efficiency is displayed as the percentage of reads with the intended 2298T>C mutation. A 15-nt PBS, 52-nt RTT, and 20-nt GPS resulted in 22% editing.

[0250] PBS lengths of 9, 11, 13, and 15-nt were tested; RTT lengths of 32, 34, 36, 52, and 56-nt were assessed; and a 20-nt GPS was included in the 52 and 56-nt RTT constructs. All of the constructs that did not include GPS resulted in under 6.3% editing, while in contrast constructs with a 52-nt RTT and 20-nt GPS gave the highest editing efficiencies with an increase from approximately 6.5% with a 9-nt PBS to 22% with a 15-nt PBS. These results indicate that GPS can significantly improve the efficiency of introducing edits as close as 23-nt from the site of the nick. Deep sequencing of the spacer's top five *in silico*-predicted off-target genomic sites was performed, and no edits introduced by transfecting the Rewriter components were detected.

[0251] Next, types of allele variants generated by the 15-nt PBS, 52-nt RTT, and 20-nt GPS construct were analyzed, and it was found that the most frequent variant contained both the 2298T>C and 2316C>A mutations (18.8%), followed by the 2316C>A mutation only (3.9%), and then the 2298T>C mutation only (3.2%) (**FIG. 25B**). **FIG. 25B** shows that the most frequent mutant allele generated by the 15-nt PBS, 52-nt RTT, 20-nt GPS construct included both the 2298T>C and 2316C>A mutations encoded in the RTT (18.8%). An additional 7.1% of reads were represented by either the 2316C>A PAM-disrupting mutation alone or the target 2298T>C mutation alone. A low frequency of adenine insertions were also detected along poly-A tracts within the target sequence. Some data for some additional aspects are shown in **FIG. 28**.

[0252] Encoding mutations in the RTT that disrupt the PAM may increase editing efficiency. Encoding mutations in the RTT that disrupt the PAM may increase editing precision. It was found that by encoding a sequence in the RTT that would disrupt the PAM site increased the efficiency of editing and decreases undesirable deletions (**FIG. 25C** and **FIG. 25G**).

[0253] There were no detectable indels generated by nSpCas9 or scaffold sequence insertion events, as may be for prime editors. Two alleles containing an adenine insertion were identified within the region that was reverse transcribed in addition to the intended 2298T>C and 2316C>A

mutations at a frequency of 0.4% and 0.2%, respectively. Each of the adenine insertions was at the end of a poly-A tract, potentially indicating that RT-mediated genome editing approaches can synthesize rare frame-shifting mutations on mononucleotide tracts of RNA templates. It was found that disrupting one of the polyA tracks in the RTT with a silent 2307A>G mutation eliminated the undesirable insertion of an adenine within that polyA track (**FIG. 25E**). It was also found that the 2298T>C editing efficiency increased to 41.6% by increasing the RTT length to 54-nt (**FIG. 25F**). Finally, it was found that not including the silent pam-disrupting 2316C>A mutation decreased the efficiency of making the 2298T>C edit 2-fold (**FIG. 25G**). Graphical data shown in any of **FIGs. 25A-25G** include a mean $\pm$ one standard deviation from three biologically independent samples; NS = not significant ($P < 0.05$; two-sided student's t-test); and ND = not detected.

[0254] As shown in **FIG. 25E** and **FIG. 30**, mutations may be encoded in the RTT to break up tracks of consecutive nucleotides (e.g. 4+ consecutive nucleotides). Undesired insertions were observed on tracks of at least 4 consecutive nucleotides containing the same base. It was considered that the reverse transcriptase was making rare insertions on these mononucleotide tracks relative to its template sequence in the RTT. It was discovered that by incorporating a mutation to break up the mononucleotide track into tracks of no more than 3 consecutive nucleotides of the same base that the undesirable insertions were no longer detected. As shown in the example in **FIG. 25E**, without encoding the silent 2307A>G mutation in the polyA track of the RTT, almost 0.5% of reads contained an undesired A insertion at position 2305. The undesirable insertion was not detected with the RTT included the silent 2307A>G mutation in the polyA track of the RTT. of the same base eliminates undesirable edits that were not encoded in the RTT

[0255] Some highlights of the Usher syndrome data include: GPS improved editing efficiency by about 4-fold, no off target effects were observed, no undesirable mutations were made to *USH2A*, and over 40% editing efficiency was achieved.

[0256] **FIG. 31A-31B** show that precise shortening of trinucleotide a repeat was achieved in an *htt* gene, demonstrating applicability of some systems and methods described herein for treating a disease such as Huntington's disease.

Discussion

[0257] An editing system such as Rewriter may comprise a targeted and efficient technology for introducing nucleotide substitutions, insertions, deletions, or complex sequence changes within approximately 70-nt of a given Cas9 target site. In addition, the ability to package Rewriter

within AAV promises to enable safe and tissue-specific delivery to treat a wide-range of genetic diseases.

[0258] Precision genome editing has traditionally relied on generating DSBs, which are in some cases genotoxic lesions that can even cause the loss of an entire chromosome. Rewriter may avoid safety concerns associated with some DSBs by only generating one single-stranded nick, generally a relatively innocuous modification. Additionally, Rewriter's deliverability and safety may not come at the cost of efficiency, as up to 64% editing was achieved, which is the highest efficiency reported to date for targeted multi-nucleotide editing in human cells without generating DSBs.

[0259] GPS may include a novel component in the Rewriter platform that may improve editing efficiency and window length by controlling the tertiary structure of the gRNA extension. GPS may relieve a constraint of requiring a PAM immediately adjacent to the site of the edit and may enable correction of multiple pathogenic mutations with a single construct. For example, the second most common *USH2A* mutation that leads to loss of vision may be 2276G>T, which may be 23-nt from the most common mutation, 2299delG. Some embodiments include use of a gRNA that is capable of treating patients with one of these mutations.

[0260] Screening of mlvRT mutants led to identification of the more efficient mlvRT14M, highlighting the potential to further optimize this component. An unbiased library of mlvRT14M mutants may be screened with the BFP to GFP conversion assay in a pooled format to improve editing. Screening a library of RTs in low dNTP concentrations, perhaps through overexpression of SAMHD1 (T592A) may identify a variant with a low enough K_M for dNTPs to obviate a possible need for VPX in order to edit non-dividing cells.

[0261] Precise editing in non-dividing cells has traditionally been a significant challenge. The results provided herein using VPX to counteract the restriction in editing caused by SAMHD1 offer a route to edit clinically relevant post-mitotic cells, such as photoreceptors and neurons, or slowly dividing cells that make up many organs. Given the identification of SAMHD1 as a potential restriction factor for editing, SAMHD1-inhibiting small molecules can be evaluated to provide a transient increase in cellular dNTP concentrations.

[0262] As the first system that, upon AAV delivery, can precisely generate targeted, complex sequence changes in the genomes of human cells without generating DSBs, Rewriter may be used to advance functional genomic studies and treat human disease.

Methods

[0263] General methods: Q5 DNA polymerase (New England Biolabs) was used for DNA amplification. DNA oligonucleotides were obtained from Integrated DNA technologies.

Plasmids were constructed by the Golden Gate assembly method. Vectors for mammalian cell experiments were purified using Plasmid Plus midiprep kits (Qiagen) or ZymoPURE miniprep kits (Zymo Research).

[0264] General mammalian cell culture: HEK293T cells (ATCC CRL-3216) were cultured and passaged in Dulbecco's modified Eagle's medium (DMEM) plus GlutaMAX (ThermoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (Gibco) and Antibiotic-Antimycotic (ThermoFisher Scientific) (DMEM+). Cells were cultured at 37°C with 5% CO₂.

[0265] Transfection: HEK293T cells were seeded on 96-well poly-d-lysine coated plates (Corning). Approximately 24 hours after seeding, media was replaced with Opti-MEM (Gibco) and each well was transfected with 0.8 µl Lipofectamine 2000 (ThermoFisher Scientific) according to the manufacturer's protocol and 400 ng of total plasmid DNA. Media was replaced with DMEM+ between 6 and 8 hours after transfection.

[0266] AAV packaging, harvest, and transduction: HEK293T cells were subjected to a triple-transfection method for production of AAV by co-transfection of three plasmids - a Rep/Cap plasmid, a helper plasmid containing adenoviral genes, and a transfer plasmid containing the cargo intended for packaging flanked by inverted terminal repeats. Transfections were performed using branched polyethylenimine (PEI) with an average molecular weight of 25,000 (Sigma 408727). Three days after transfection, cells were harvested and purified using the AAVpro Purification Kit Maxi (Takara 6666). Titers of purified AAV stocks were determined by qPCR on a CFX96 Real-Time System (Bio-Rad) using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). BFP-expressing HEK293 cells were co-transduced with equal numbers of AAV-A and AAV-B viral particles, and editing of BFP-to-GFP was determined 96-120 hours after transduction by flow cytometry.

[0267] Flow cytometry: 48 hours after transfection, media was removed and cells were detached with 0.05% Trypsin/EDTA (Gibco). Trypsin was neutralized with DMEM+ and suspended cells were placed in round-bottom 96-well plates. An Attune NxT flow cytometer (ThermoFisher Scientific) was used to analyze the fluorescence of 30,000 cells per well.

[0268] High-throughput genomic DNA sequencing: Genomic sites of interest were amplified from genomic DNA samples and sequenced on an Illumina MiSeq. Amplification primers comprising Illumina forward and reverse adapters were used for a first round of PCR (PCR 1) to amplify a genomic region of interest. PCR 1 reactions (25 µl) were performed with 0.5 µM of each forward and reverse primer, 1 µl genomic DNA extract and 12.5 µl Phusion U Green Multiplex PCR Master Mix. PCR reactions were carried out as follows: 98°C for 2 min, then 30 cycles of [98°C for 10 s, 61°C for 20 s, and 72°C for 30 s], followed by a final 72°C extension

for 2 min. Unique Illumina barcoding primer pairs were added to each sample in a secondary PCR reaction (PCR 2). Specifically, 25 μ l of a given PCR 2 reaction contained 0.5 μ M of each unique forward and reverse Illumina barcoding primer pair, 1 μ l unpurified PCR 1 reaction mixture, and 12.5 μ l of Phusion U Green Multiplex PCR 2 Master Mix. The barcoding PCR 2 reactions were carried out as follows: 98°C for 2 min, then 12 cycles of [98°C for 10 s, 61°C for 20 s, and 72°C for 30 s], followed by a final 72°C extension for 2 min. PCR products were evaluated analytically by electrophoresis in a 1.5% agarose gel. PCR 2 products (pooled by common amplicons) were purified by electrophoresis with a 1.5% agarose gel using a QIAquick Gel Extraction Kit (Qiagen), eluting with 40 μ l water. DNA concentration was measured by fluorometric quantification (Qubit, ThermoFisher Scientific) or qPCR (KAPA Library Quantification Kit-Illumina, KAPA Biosystems) and sequenced on an Illumina MiSeq instrument according to the manufacturer's protocols. Sequencing reads were demultiplexed using MiSeq Reporter (Illumina).

[0269] Alignment of amplicon sequences to a reference sequence was performed using CRISPResso2⁴³. For all prime editing yield quantification, prime editing efficiency was calculated as: percentage of (number of reads with the desired edit that do not contain indels)/(number of total reads). For quantification of point mutation editing, CRISPResso2 was run in standard mode with “discard_indel_reads” on. Prime editing for installation of point mutations was then explicitly calculated as: (frequency of specified point mutation in non-discarded reads) $\sim$ (number of non-discarded reads)/(total reads). For insertion or deletion edits, CRISPResso2 was run in HDR mode using the desired allele as the expected allele (e flag), and with “discard_indel_reads” on. Editing yield was calculated as: (number of HDR-aligned reads)/(total reads). Indel yields were calculated as: (number of indel comprising reads)/(total reads).

[0270] 36 hours after transfection, cells were detached with 0.05% trypsin, spun down, washed with PBS, and spun down again. Cell pellets were resuspended in 10 μ l of QuickExtract (Lucigen) and incubated at 65°C for 6 minutes. Samples were then vortexed for 15 seconds and incubated at 98°C for 2 minutes. 10 μ l of nuclease-free water was added to each sample. 4 μ l of sample was used as a template for PCR with Q5 polymerase and primers that contain Illumina adapters that were designed to amplify the genomic region of interest. Samples were then treated with Exo-CIP (NEB) at 37°C for 1 hour. DNA concentration was measured with Qubit (ThermoFisher Scientific) and samples were sent to Genewiz for sequencing using the Amplicon-EZ service. PE-Analyzer (<http://www.rgenome.net/pe-analyzer>) was used to analyze high-throughput sequencing data. The highest frequency variant in control samples that were not

transduced with any genome editing components was set as the detection threshold and any variant below this frequency was discarded. The efficiency of installing the 2298T>C mutation was explicitly calculated as (number of reads containing only the 2298T>C mutation + number of reads containing only the 2298T>C and 2316C>A mutations)/(total number of reads of alleles that were present at a frequency above the detection threshold).

[0271] Unless otherwise defined, all technical terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. As used in this specification and the appended claims, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise. Any reference to “or” herein is intended to encompass “and/or” unless otherwise stated.

[0272] Whenever the term “at least,” “greater than,” or “greater than or equal to” precedes the first numerical value in a series of two or more numerical values, the term “at least,” “greater than” or “greater than or equal to” applies to each of the numerical values in that series of numerical values. For example, greater than or equal to 1, 2, or 3 is equivalent to greater than or equal to 1, greater than or equal to 2, or greater than or equal to 3.

[0273] Whenever the term “no more than,” “less than,” “less than or equal to,” or “at most” precedes the first numerical value in a series of two or more numerical values, the term “no more than,” “less than” or “less than or equal to,” or “at most” applies to each of the numerical values in that series of numerical values. For example, less than or equal to 3, 2, or 1 is equivalent to less than or equal to 3, less than or equal to 2, or less than or equal to 1.

[0274] Where values are described as ranges, it will be understood that such disclosure includes the disclosure of all possible sub-ranges within such ranges, as well as specific numerical values that fall within such ranges irrespective of whether a specific numerical value or specific sub-range is expressly stated.

EXAMPLES

[0275] The following examples are illustrative and non-limiting to the scope of the devices, methods, systems, and kits described herein.

EXAMPLE 1

Genome Editing Efficiency Assays

[0276] This example describes genome editing efficiency assays. Precision editing rates of genome editing constructs were determined by measuring the frequency of editing a blue fluorescent protein (BFP) gene to produce green fluorescent protein (GFP). Specifically, 30,000 HEK293T cells with a genomically-integrated BFP gene were seeded in 96-well poly-d-lysine-

treated plates in DMEM containing 10% fetal bovine serum (FBS). After 12-24 hours, media was replaced with opti-mem media. Lipofectamine 2000 was used to transfect plasmids encoding genome editing components. 25 microliters of opti-mem containing a total of 400 nanograms of plasmid DNA was added to 25 microliters of opti-mem containing 0.8 microliters of Lipofectamine 2000. After 20 minutes, the 50 microliter mixture was added drop-wise to the well containing cells. After 6 hours, media was replaced with DMEM containing 10% FBS. GFP and BFP levels were measured 36-60 hours later using an Attune NxT flow cytometer.

EXAMPLE 2

Editing Efficiency of a Split nCas9 Reverse Transcriptase Construct

[0277] This example describes the editing efficiency of a split nCas9 reverse transcriptase construct. Plasmids encoding either a fused nCas9-mlvRT or a split nCas9-RT and a gRNA were prepared and transfected as described in **EXAMPLE 1**. Editing efficiency of each construct was measured. **FIG. 1** shows the editing efficiency of a fused Cas9 nickase (nCas9) reverse transcriptase (RT) construct (“nCas9-mlvRT”) comprising an nCas9 and a Moloney leukemia virus RT (mlvRT), and a split nCas9-LZ1 and LZ2-mlvRT construct (“mlvRT Split Stitch”). Split Stitch may be referred to as Rewriter (e.g. RWb1), or vice versa. In some cases, a Split Stitch may include a Rewriter (e.g. RWb1) or a Rewriter component. In some cases, a Rewriter may include a Split Stitch or a Split Stitch component. mlvRT Split Stitch may be an example of a component of Rewriter (e.g. RWb1). The split nCas9-LZ1 and LZ2-mlvRT construct comprises a nCas9-LZ1 (SEQ ID NO: 1, NLS-SpCas9(H840A)-NLS-EE12RR345L(leucine zipper)) and a LZ2-mlvRT (SEQ ID NO: 2, RR12EE345L(leucine zipper)-mlvRTv(nCas9-mlvRT(D200N, L603W, T306K, W313F, T330P)-NLS) on discrete polypeptide chains. The nCas9-LZ1 comprises a SpCas9 (SEQ ID NO: 32) and a C-terminal leucine zipper (SEQ ID NO: 23) that heterodimerizes with the LZ2-mlvRT comprising a mlvRT (SEQ ID NO: 13) and an N-terminal leucine zipper (SEQ ID NO: 24) through the leucine zippers. Schematics of the nCas9-mlvRT constructs are provided at the top of the figure. The split nCas9-LZ1 and LZ2-mlvRT construct showed improved editing efficiency (about 38% efficiency) nearly double that of the fused nCas9-RT construct (about 21% efficiency).

[0278] **FIG. 11A** shows domain arrangements of a prime editor 2 system (“PE2,” top), a split prime editor 2 system (“split PE2,” middle), and a split stitch construct with two leucine zippers (“Split Stitch,” bottom). On the right is a structural schematic of the Split Stitch construct comprising a Cas9 nickase (nCas9) and a reverse transcriptase (RT) linked by two leucine zippers (LZ1 and LZ2) complexed with a guide nucleic acid. The Split Stitch split nCas9-LZ1 and LZ2-mlvRT construct comprises a nCas9-LZ1 (SEQ ID NO: 1, NLS-SpCas9(H840A)-NLS-

EE12RR345L(leucine zipper)) and a LZ2-mlvRT (SEQ ID NO: 2, RR12EE345L(leucine zipper)-mlvRTv(nCas9-mlvRT(D200N, L603W, T306K, W313F, T330P)-NLS) on discrete polypeptide chains.

EXAMPLE 3

Effect of gRNA Hairpin Inserts on Reverse Transcriptase Recruitment

[0279] This example describes the effect of gRNA hairpin inserts on editing efficiency. Plasmids encoding either a fused nCas9-mlvRT or a split nCas9 and mcp-RT and a gRNA were prepared and transfected as described in **EXAMPLE 1**. Editing efficiency of the fused nCas9-RT was measured in the presence of pegRNA. Editing efficiency of the split nCas9-RT was measured in the presence of three different gRNAs either with hairpins embedded in the scaffold (gRNA 2.0) or with hairpins of varying lengths (1xlongMS2, 1xshortMS2, or 2xshortMS2) positioned after the scaffold. **FIG. 2** shows the editing efficiency of a fused nCas9-RT construct (“nCas9-mlvRT”) and a split nCas9 and mcp-mlvRT construct (“mcp-mlvRTv”) comprising an nCas9 and a mcp peptide fused to reverse transcriptase (SEQ ID NO: 27). The mcp peptide interacts with MS2 RNA hairpins. Efficiency of the split nCas9 and mcp-mlvRT construct was tested with different guide RNA (gRNA) constructs including gRNA 2.0 (SEQ ID NO: 31), a gRNA with a long MS2 hairpin (SEQ ID NO: 28), “gRNA-1xlongMS2”), a gRNA with a short MS2 hairpin (SEQ ID NO: 29, “gRNA-1xshortMS2”), or a gRNA with two short MS2 hairpins (SEQ ID NO: 30, “gRNA-2xshortMS2”). The gRNA with the 1xlongMS2 hairpin and the gRNA with the 1xshortMS2 hairpin showed improved editing efficiency over gRNA 2.0.

[0280] **FIG. 11B** shows the editing efficiency of the constructs illustrated in **FIG. 11A** with different gRNAs. Editing efficiency was measured as a percentage of cells that were edited to convert a BFP to a GFP (% GFP+). Editing efficiency was tested with different guide RNA (gRNA) constructs including gRNA 2.0 (SEQ ID NO: 31), a gRNA with a long MS2 hairpin (SEQ ID NO: 28), “gRNA-1xlongMS2”), a gRNA with a short MS2 hairpin (SEQ ID NO: 29, “gRNA-1xshortMS2”), or a gRNA with two short MS2 hairpins (SEQ ID NO: 30, “gRNA-2xshortMS2”). The Split Stitch construct (RWb1 in this instance) showed improved editing efficiency over the prime editor 2 (PE2) construct.

EXAMPLE 4

Split nCas9-RT Construct with Increased Reverse Transcriptase Processivity

[0281] This example describes a split nCas9-RT construct with increased reverse transcriptase processivity. The split nCas9-RT construct described in **EXAMPLE 2** was further engineered to increase the processivity of the reverse transcriptase polymerase function. The reverse

transcriptases with increased processivity was able to catalyze the formation of more sequential phosphodiester bonds in a single binding event than the reverse transcriptase provided in **EXAMPLE 2**. The increased processivity facilitated the reverse transcription of longer template sequences and may enable editing of longer sequences at a target site of a genome. The editing efficiency of three split nCas9-RT constructs with reverse transcriptases having increased processivity were tested.

[0282] FIG. 3 shows the editing efficiency of different split nCas9-RT constructs comprising modified reverse transcriptases with increased transcriptional processivity. Constructs comprising nCas9 and reverse transcriptases from either *geobacillus stereothermophilus* (Gsl-IICRT, SEQ ID NO: 3), *Eubacterium Rectale* (ErRT, SEQ ID NO: 4), and amino acids 116-1016 from the R2 polyprotein (R2(116-1016), SEQ ID NO: 7) were tested. A schematic of the Gsl-IICRT reverse transcriptase (“StitchRT”) is shown compared to the mlvRT reverse transcriptase used in **FIG. 1** and **FIG. 2**. Split Stitch R2(116-1016) showed the highest editing efficiency of the three split nCas9 and RT constructs comprising modified reverse transcriptases with increased transcriptional processivity tested.

EXAMPLE 5

gRNAs for Increased Editing Efficiency at Single-Strand Breaks

[0283] This example describes gRNAs for increased editing efficiency at single-strand breaks. gRNAs were designed to increase efficiency of editing at a single-strand break by incorporating a second strand primer at the 3' end of the gRNA. The second strand primer primed the synthesis of the second strand using a newly synthesized first strand as a template. Priming of second strand synthesis facilitated the insertion of the synthesized sequence into the site of a single-strand break without formation of a double-strand break. Formation of double-strand breaks may increase the rate of formation of undesired products.

[0284] FIG. 4A illustrates a method for genome editing using an engineered gRNA of the present disclosure (“Stitch Guide”). In some cases, a Stitch Guide may include a Rewriter (e.g. Rewriter 3.0, Rewriter 3.1, or Rewriter 3.2) or a Rewriter component. In some cases, a Rewriter may include a Stitch Guide or a Stitch Guide component. A nCas9-RT construct complexed with a gRNA is recruited to a target site of a target nucleic acid by hybridization of a spacer of the gRNA to the target site. The nCas9 nicks a strand of a target nucleic acid at a target site. A first strand primer binding site of the gRNA hybridizes to a flap 5' of the nick. The RT polymerizes from the 3' end of the flap using a reverse transcriptase template region of the gRNA as a template. A second strand primer (“2nd strand primer”) at the 3' end of the gRNA hybridizes to the 3' end of the newly synthesized DNA strand. The 4-300 bp second strand primer region acts

as an RNA primer for synthesis of a second DNA strand. The RT polymerizes from the 3' end of the gRNA using the newly synthesized DNA strand as a template. A ribozyme on the 3' end of the gRNA cleaves the gRNA 3' of the second strand primer sequence. The newly synthesized double stranded DNA may be incorporated into the target nucleic acid at the site of the nick.

[0285] FIG. 4B shows the editing efficiency of a nCas9-RT construct using a pegRNA gRNA or a Stitch Guide gRNA. Schematics of the pegRNA and the Stitch Guide gRNA are shown at left. The inclusion of the second strand primer in the Stitch Guide gRNA improved the editing efficiency relative to the pegRNA lacking the second strand primer. The fused nCas9-mlvRTv construct was used in this assay.

[0286] In a second assay, the effect of second strand primer length and binding site on editing efficiency of split nCas9-RT constructs was tested. The length of the second strand primer was varied as well as the binding position of the second strand primer relative to the position of the single strand break. **FIG. 5A** shows the editing efficiency of a fused nCas9-RT construct ("nCas9-mlvRTv") with different gRNAs comprising second strand primers (SSPs) 20 nucleotides (nt), 40 nt, or 60 nt in length positioned either 6 nt, 36, nt, or 55 nt 3' of the 5' end of the first strand primer binding site ("nt from nick"). A gRNA lacking a second strand primer was tested as a control. All gRNA sequences comprise an hdv ribozyme (SEQ ID NO: 25). **FIG. 5B** shows the editing efficiency of a nCas9-RT ("nCas9-R2(116-1016)") with different gRNAs comprising second strand primers (SSPs) 20 nucleotides (nt), 40 nt, or 60 nt in length that positioned either 6 nt, 36, nt, or 55 nt 3' of the 5' end of the first strand primer binding site ("nt from nick"). A gRNA lacking a second strand primer was tested as a control. With both the nCas9-mlvRTv and the nCas9 and R2 constructs, the gRNAs with shorter (e.g., 20 nt) second strand primers showed improved editing efficiency as compared to the other gRNAs with longer second strand primers.

EXAMPLE 6

Dual Guide Systems for Improved Editing

[0287] This example describes dual guide systems for improved editing. Dual guide systems comprising two gRNAs targeting two target sites on opposite strands in close proximity are introduced into a cell. Each gRNA recruits a nCas9-RT contract to the respective target site, facilitating a single strand break at each target site. The two gRNAs are fused for improved delivery and to ensure co-localization to the two target sites.

[0288] FIG. 6 illustrates four schemes of genome editing using a two gRNA system with a nCas9-RT. In a two single guide system in which the two guides each generate an edited strand (top left), each gRNA binds to a different nCas9 and the two gRNAs each comprise a reverse

transcriptase template region and a primer binding site (PBS) region. In a two single guide system in which the second guide nicks the opposite strand (top right), each gRNA binds to a different nCas9 and only one of the gRNAs comprise a reverse transcriptase template region and a primer binding site (PBS) region. In a dual guide complex system in which the two guides each comprise a reverse transcriptase template region and a primer binding site (PBS) region (bottom left), the spacer of the first gRNA binds the first nCas9, the spacer of the second gRNA binds the second nCas9, the scaffold of the first gRNA binds the second nCas9, and the scaffold of the second gRNA binds the first nCas9; and the two gRNAs each comprise a reverse transcriptase template region and a PBS region. In a dual guide complex system in which the second guide nicks the opposite strand (bottom right), the spacer of the first gRNA binds the first nCas9, the spacer of the second gRNA binds the second nCas9, the scaffold of the first gRNA binds the second nCas9, and the scaffold of the second gRNA binds the first nCas9; and only one of the gRNAs comprise a reverse transcriptase template region.

[0289] FIG. 7 illustrates a method for increasing the efficiency of gene editing. A two single guide system in which the second guide nicks the opposite strand or a dual guide complex system in which the second guide nicks the opposite strand, the nick on the opposite strand facilitates incorporation of the newly synthesized DNA into the target nucleic acid. The second guide generates a flap that is reverse complementary to a region in the of first newly synthesized strand. The first synthesized strand acts as template for second strand synthesis.

[0290] Editing efficiency of a two gRNA system was measured by introducing a stop codon into a target nucleic acid encoding a blue fluorescent protein. Assays were performed as described in **EXAMPLE 1** except that successful editing was identified by a lack of BFP fluorescence. **FIG. 10** shows the editing efficiency of a two gRNA system as illustrated in **FIG. 7**. Editing efficiency, as measured by percent of cells negative for BFP (“% BFP-”), was measured for cells only (no gRNA), single gRNAs (gRNA 2 which lacks a 3’ extension, gRNA 1 without a stub, and gRNA 1 with a stub), and two gRNAs (gRNA 1 without a stub plus gRNA 2 and gRNA 1 with a stub and gRNA 2). The two gRNA systems increased editing efficiency as compared to the single gRNA systems. Presence of a stub in gRNA 1 in the two gRNA system increased editing efficiency compared to the two gRNA system lacking a stub in gRNA 1.

EXAMPLE 7

gRNA Velcro for Improved Editing

[0291] This example describes gRNA Velcro for improved editing. A gRNA comprising a Velcro region improved the efficiency of strand formation by facilitating an interaction between the gRNA and a flap formed 5’ of the nick in the target nucleic acid. The Velcro region was

positioned either 5' of the reverse transcriptase template region or 3' of the first strand primer binding site. The gRNA Velcro insertion was compatible with the single guide systems provided in **EXAMPLE 2 – EXAMPLE 5** or dual guide systems provided in **EXAMPLE 6**.

[0292] FIG. 8A illustrates a gRNA comprising a Velcro region to accelerate the rate of hybridization of the primer binding site and the flap by creating regions of reverse complementation within the 3' extended guide RNA. The Velcro region comprises 5 to 200 nucleotides positioned 5' of the reverse transcriptase template region that are reverse complementary to the region of the gRNA 5' of the first strand primer binding site. **FIG. 8B** illustrates a gRNA comprising a Velcro region to accelerate the rate of hybridization of the primer binding site and the flap by creating regions of reverse complementation within the 3' extended guide RNA. The Velcro region comprises 5 to 100 nucleotides positioned 3' of the first strand primer binding site that are reverse complementary to the region 5' of the reverse transcriptase template region.

[0293] FIG. 12A illustrates gRNA constructs either without (left) or with (middle and right) a Velcro region to accelerate the rate of hybridization of the primer binding site (PBS) to a flap of a target nucleic acid. In a V1 arrangement, the Velcro region may be positioned at or near the 5' end of the gRNA and may hybridize to a region of the gRNA 5' of the primer binding site ("Velcro V1," middle). In a V2 arrangement, the Velcro region may be positioned 3' of the primer binding site and may hybridize to a region at or near the 5' end of the gRNA ("Velcro V2," right).

[0294] FIG. 12B illustrates predicted three-dimensional structures of the gRNA constructs provided in **FIG. 12A**. A gRNA lacking a Velcro region is shown in the left. gRNAs comprising a Velcro V1 region or a Velcro V2 region are shown in the middle and right panels, respectively.

[0295] FIG. 9A shows the editing efficiency of a nCas9-LZ1 and LZ2-mlvRTv construct with the gRNA constructs comprising a Velcro region, as illustrated in **FIG. 8A** and **FIG. 8B**. Editing efficiency was compared using a gRNA lacking a Velcro region ("no Velcro"), a 15 nt Velcro region positioned 5' of the reverse transcriptase template region ("V1," as illustrated in **FIG. 8A**) with a gap length of 1, 5, or 10 nts, or a Velcro region positioned 3' of the first strand primer binding site ("V2," as illustrated in **FIG. 8B**) of either 10 or 20 nt in length. The gRNA contained a 107 nucleotide RT template, and a 13 nucleotide primer binding site. Editing was performed such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA. **FIG. 9B** shows the editing efficiency of a nCas9-LZ1 and LZ2-R2(116-1016) construct with the gRNA constructs comprising a Velcro region, as illustrated in **FIG. 8A** and **FIG. 8B**. Editing efficiency was compared using a gRNA lacking a Velcro region ("no Velcro"), a 15 nt

Velcro region positioned 5' of the reverse transcriptase template region ("V1," as illustrated in **FIG. 8A**) with a gap length of 1, 5, or 10 nts between the end of the Velcro binding site and the beginning of the primer binding site, or a Velcro region positioned 3' of the first strand primer binding site ("V2," as illustrated in **FIG. 8B**) of either 10 or 20 nt in length. With both nCas9-RT constructs, certain gRNAs comprising Velcro regions increased editing efficiency. In particular, the V1 Velcro gRNA with the 1 nt gap and the 20 nt Velcro gRNA improved editing efficiency in the nCas9-mlvRTv construct, and the V1 Velcro gRNA with the 10 nt gap and the 20 nt V2 Velcro gRNA improved editing efficiency in the nCas9-R2 construct.

[0296] FIG. 12C shows editing efficiency of a gRNA with a 129 nucleotide RT template and a 13 nucleotide primer binding site and a 20 nucleotide Velcro region. Editing was performed such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT. Editing efficiency was compared for the original gRNA ("original coding") or a gRNA recoded with silent mutations in the RT template to remove secondary structure ("recoded"). Removal of secondary structure using silent mutations improved editing efficiency relative to the original RT template. Additionally, use of gRNA comprising a Velcro region allowed efficient editing at a distance of 65 nucleotides from the nicking site.

[0297] FIG. 12D shows editing efficiency of gRNAs with different lengths of Velcro sequences. Each gRNA contained, in order from 5' to 3', a RT template, a primer binding site, and a Velcro region, as shown in the schematic on the left. Editing efficiency was measured as the percent of cells that were GFP positive (% GFP+). gRNAs had a 129 nucleotide RT template, a 13 nucleotide primer binding site. Editing was performed such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT. The gRNA with a 20 nucleotide Velcro region positioned at the 3' end with no gap showed higher editing efficiency than the other gRNAs tested.

EXAMPLE 8

Co-Delivery of Protective Complexes for Improved Editing System Delivery

[0298] This example describes co-expression of protective complexes for improved delivery of the editing systems provided herein. A nCas9-RT constructs provided herein and a gRNA provided herein are delivered to a cell. The nCas9-RT and the gRNA are co-expressed with an open reading frame sequence encoding protective protein complexes. The protective protein complexes are expressed in the cell, preventing degradation or deamination of the gRNA, thereby improving delivery of the editing system. The open reading frame sequence is a Human Orf1p (SEQ ID NO: 38) or a Murine Orf1p (SEQ ID NO: 39).

EXAMPLE 9**Improved Editing Efficiency with gRNAs with a Velcro Region and a Second Strand Primer**

[0299] This example describes improved editing efficiency with gRNAs with a Velcro region and a second strand primer. gRNAs were designed to increase efficiency of editing at a single-strand break by incorporating a second strand primer at the 3' end of the gRNA and a Velcro region 5' of the second strand primer. The second strand primer primed the synthesis of the second strand using a newly synthesized first strand as a template. Priming of second strand synthesis facilitated the insertion of the synthesized sequence into the site of a single-strand break without formation of a double-strand break. The Velcro region improved the efficiency of strand formation by facilitating an interaction between the gRNA and a flap formed 5' of the nick in the target nucleic acid.

[0300] FIG. 13A illustrates schematics of a pegRNA and a Stitch gRNA comprising a Velcro region and a 2nd strand primer (top) and a method of genome editing using a Stitch gRNA (bottom). A nCas9-RT construct complexed with a gRNA is recruited to a target site of a target nucleic acid by hybridization of a spacer of the gRNA to the target site. The nCas9 nicks a strand of a target nucleic acid at a target site. A first strand primer binding site of the gRNA hybridizes to a flap 5' of the nick. The RT polymerizes from the 3' end of the flap using a reverse transcriptase template region of the gRNA as a template. A second strand primer ("2nd strand primer") at the 3' end of the gRNA hybridizes to the 3' end of the newly synthesized DNA strand. The 4-200 bp second strand primer region acts as an RNA primer for synthesis of a second DNA strand. The RT polymerizes from the 3' end of the gRNA using the newly synthesized DNA strand as a template. A ribozyme on the 3' end of the gRNA cleaves the gRNA 3' of the second strand primer sequence. The newly synthesized double stranded DNA may be incorporated into the target nucleic acid at the site of the nick.

[0301] FIG. 13B shows editing efficiency of gRNAs second strand primers (SSPs) of varying lengths and that hybridize at varying distances from the nicking site. Second strand primers 20, 40, or 60 nucleotides (nt) long positioned 6, 36, or 55 nucleotides from the nick were tested. Editing efficiency was measured as the percent of cells that were GFP positive (% GFP+).

[0302] FIG. 13C shows editing efficiency of gRNAs without a Velcro region or a second strand primer ("no velcro, no SSP"), with a 19 nucleotide Velcro region ("19 nt velcro"), or with both a 19 nucleotide Velcro region and a 20 nucleotide second strand primer ("19 nt velcro, 20 nt SSP"). Editing efficiency was increased when using a gRNA containing both a Velcro region and a second strand primer as compared to a gRNA lacking a Velcro region and a second strand

primer or a gRNA containing a Velcro region but no second strand primer. The editing efficiency achieved using the a gRNA containing both a Velcro region and a second strand primer was 54% from a single nick, which was higher than the predicted limit of 50% for editing efficiency from a single nick.

EXAMPLE 10

Reverse Transcriptase Protein Engineering to Increase Editing Efficiency

[0303] This example describes reverse transcriptase protein engineering to increase editing efficiency. Point mutations were made in an mlvRT construct to improve editing efficiency. Editing efficiency was measured using the mutated constructs.

[0304] FIG. 14A shows the results of a screen for mutations in a mlvRT reverse transcriptase and their effect on editing efficiency. Mutations were made in a reference mlvRT construct containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing a Y8H, P51L, S56A, S67R, E69K, Q84A, F155Y, T197A, H204R, T246E, N249D, E286R, Q291I, R301L, E302K, F309N, M320L, L435G, D524A, D524G, D524N, E562D, K571R, D583N, Y586S, H594Q, H638G, D653N, T664N, or L671P single point mutation (SEQ ID NO: 41 – SEQ ID NO: 70, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 1 nucleotide gap, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA. Mutants containing Q84A (SEQ ID NO: 46), T197A (SEQ ID NO: 48), D653N (SEQ ID NO: 68), T664N (SEQ ID NO: 69), or L671P (SEQ ID NO: 70) showed significantly increased editing efficiency compared to SEQ ID NO: 40.

[0305] FIG. 14B shows the results of a screen for combinations of mutations in a mlvRT reverse transcriptase and their effect on editing efficiency. Mutations were made in a reference mlvRT construct containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing T197A and D653N; T197A and T664N; T197A and L671P; T197A, D653N, T664N and L671P; or P51L, S67R, T197A, H204R, L435G, D524A, D653N, T664N and L671P (SEQ ID NO: 71 – SEQ ID NO: 75, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an

85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA. The construct containing P51L, S67R, T197A, H204R, L435G, D524A, D653N, T664N and L671P point mutations (SEQ ID NO: 75) showed the highest editing efficiency of the constructs tested.

[0306] FIG. 15C shows the editing efficiency of mlvRT reverse transcriptase constructs.

Mutations were made in a reference mlvRT construct containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing V223A; V223M; Q221R and V223A; or Q221R and V223M (SEQ ID NO: 77 – SEQ ID NO: 80, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with a 129 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT. The V223A point mutation increased editing efficiency at a distance relative to the nicking site.

EXAMPLE 11

Increasing Availability of dNTPs to Increase Editing Efficiency

[0307] This example describes methods for increasing availability of dNTPs in a cell to increase editing efficiency of a Cas9-RT construct. One factor that may contribute to low editing efficiency in cells is limited availability of dNTPs. **FIG. 15A** illustrates a method of increasing availability of dNTPs in a cell to increase editing efficiency. In non-dividing cells lacking CDK1, unphosphorylated SAMHD1 cleaves dNTPs, decreasing the available dNTPs in the cell. In dividing cells, CDK1 phosphorylates SAMHD1, preventing SAMHD1 from cleaving dNTPs and leading to increased availability of dNTPs in the cell. A single point mutation in SAMHD1 (T592A) prevents phosphorylation of SAMHD1 by CDK1, resulting in a constitutively active SAMHD1 and a low availability of dNTPs in the cell. The T592A mutant SAMHD1 was used to induce a low dNTP environment in the assay shown in **FIG. 15B**, **FIG. 15D**, and **FIG. 15E**. Addition of Vpx inhibits SAMHD1, leading to increased availability of dNTPs in the cell. To test the effect of low cellular dNTPs on editing efficiency, constitutively active SAMHD1 was co-expressed with Cas9-RT constructs, and editing efficiency was measured.

[0308] FIG. 15B shows the editing efficiency of mlvRT reverse transcriptase constructs in the presence or absence of a constitutively active SAMHD1 (SAMHD1 (T592A)) to decrease availability of dNTPs in the cell. Mutations were made in a reference mlvRT construct

containing five point mutations (D200N, I603W, T330P, T306K, and W313F, SEQ ID NO: 40). Amino acid residues are counted relative to an mlvRT construct lacking an N-terminal methionine (e.g., SEQ ID NO: 14). mlvRT constructs containing Q221R; V223A; V223M; Q221R and V223A; or Q221R and V223M (SEQ ID NO: 76 – SEQ ID NO: 80, respectively) relative to SEQ ID NO: 40 were tested. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA. Expression of the constitutively active SAMHD1 decreased editing efficiency of all tested constructs.

[0309] To rescue editing efficiency in cells expressing constitutively active SAMHD1, a Vpx peptide (SEQ ID NO: 82) was also expressed in the cells. **FIG. 15D** shows the editing efficiency of a mlvRT reverse transcriptase in the presence or absence of a constitutively active SAMHD1 (SAMHD1 (T592A)) to decrease availability of dNTPs in the cell and with or without Vpx (SEQ ID NO: 82) to inhibit SAMHD1. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0310] **FIG. 15E** shows the editing efficiency of a mlvRT reverse transcriptase in the presence or absence of a constitutively active SAMHD1 (SAMHD1 (T592A)) to decrease availability of dNTPs in the cell and with or without Vpx (SEQ ID NO: 82) to inhibit SAMHD1. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with a 129 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT. Expression of Vpx in the cell increased editing efficiency both in cells expressing constitutively active SAMHD1 and in cells not expressing constitutively active SAMHD1. Additionally, Vpx increased editing efficiency at sites a short distance from the nicking site (**FIG. 15D**) and at a long distance from the nicking site (**FIG. 15E**).

EXAMPLE 12

Inteins for Cellular Expression of Split Cas9 Constructs

[0311] This example describes using inteins for cellular expression of Cas9 constructs. In a first assay, a screen of nicking Cas9 (nCas9) point mutations was performed to identify positions in

the C-terminal portion of the nCas9 that were conducive to substitution of a cysteine residue. Cysteine point mutations were screened in the context of a nCas9-RT construct linked via a leucine zipper. Cysteines were inserted into the C-terminal portion of the nCas9 at different points to generate constructs with a cysteine residue positioned toward the middle of the nCas9 and reverse transcriptase combined sequence. Cysteine residues were positioned such that each of the portion of the Cas9 protein from the N-terminus up to the inserted cysteine and the portion of the Cas9 protein from and including the inserted cysteine to the C-terminus plus the reverse transcriptase were small enough to fit in an AAV vector when expressed as an intein fusion. Editing efficiency of the leucine zipper linked nCas9-RT cysteine mutants was compared.

[0312] FIG. 16 shows editing efficiency of Cas9 constructs modified for nicking activity and linked to a reverse transcriptase through a leucine zipper. *S. Pyogenes* Cas9 (“SpCas9”) constructs contained an H840A mutation to produce a Cas9 nickase (nCas9). Cysteine residues were introduced into the Cas9 nickase at either D1079C, S1173C, or D1180C to enable splitting of the Cas9 into a split intein Cas9 (iCas9) for expression as extein-intein fusions. Leucine zipper Cas9 constructs containing H840A and D1079C (SEQ ID NO: 85 with a leucine zipper), H840A and S1173C (SEQ ID NO: 86 with a leucine zipper), or H840A and D1180C (SEQ ID NO: 87 with a leucine zipper) point mutations and linked to mlvRT5M (SEQ ID NO: 40 with a leucine zipper) were tested. A Cas9 nickase that contained the H840A mutation but no additional cysteine (SEQ ID NO: 84 with a leucine zipper) linked to mlvRT5M (SEQ ID NO: 40 with a leucine zipper) was used as a control. Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3’ of the nick, was mutated to CATA.

[0313] In a second assay, the cysteine point mutations identified in the first assay were utilized to generate split intein Cas9 constructs. The editing efficiency of an nCas9-RT fusion comprising the identified S1173C mutation expressed as two extein-intein fusions was tested. The nCas9-RT fusion contained an nCas9 with the S1173C point mutation (SEQ ID NO: 86) fused to mlvRT5M (SEQ ID NO: 40). The first segment of the fusion protein was expressed as a first intein fusion nCas9(1-1172) – Npu N intein (SEQ ID NO: 90) in a first plasmid vector and the second segment of the fusion protein was expressed as a second intein fusion Npu C intein – nCas9(1173-1368 with S1173C) – mlvRT5M (SEQ ID NO: 91) expressed in a second plasmid vector. Autocatalytic activity of the intein domains fused the nCas9(1-1172) extein to the nCas9(1173-1368 with S1173C) – mlvRT5M extein and excised the Npu N (SEQ ID NO: 88)

and Npu C (SEQ ID NO: 89) inteins to form the fused nCas9(S1173C) – mlvRT5M construct (SEQ ID NO: 92). Editing efficiency of the split intein nCas9 construct was tested at two positions relative to the nicking site.

[0314] FIG. 17A shows the editing efficiency of a split intein Cas9 (iCas9) S1173C construct modified for nicking activity, fused to a reverse transcriptase, and expressed as two extein-intein fusion proteins. The N-terminal region of the nCas9-RT construct was expressed as nCas9(1-1172) – Npu N intein (SEQ ID NO: 90) and the C-terminal region of the nCas9-RT construct was expressed as Npu C intein – nCas9(1173-1368 with S1173C) – mlvRT5M (SEQ ID NO: 91). Editing efficiency of the split intein Cas9-RT construct (right bar) was compared to a leucine zipper split Cas9 construct (SEQ ID NO: 1 and SEQ ID NO: 2, left bar). Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

[0315] FIG. 17B shows the editing efficiency of a split intein Cas9 (iCas9) S1173C construct modified for nicking activity, fused to a reverse transcriptase, and expressed as two extein-intein fusion proteins. The N-terminal region of the nCas9-RT construct was expressed as nCas9(1-1172) – Npu N intein (SEQ ID NO: 90) and the C-terminal region of the nCas9-RT construct was expressed as Npu C intein – nCas9(1173-1368 with S1173C) – mlvRT5M (SEQ ID NO: 91). Editing efficiency of the split intein Cas9-RT construct (right bar) was compared to a leucine zipper split Cas9 construct (SEQ ID NO: 1 and SEQ ID NO: 2, left bar). Editing efficiency was measured as a percent of cells that were GFP positive (% GFP+). Editing was performed using a gRNA with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, and a 20 nucleotide second strand primer to edit a site such that an ATG sequence, starting 65 nucleotides 3' of the nick, was mutated to CAT.

[0316] The results indicated that the split intein Cas9-RT fusion construct showed robust editing efficiency compared to the leucine zipper split Cas9 constructs.

EXAMPLE 13

3' Modifications of gRNAs for Improved Editing Efficiency

[0317] This example describes 3' modifications of gRNAs for improved editing efficiency. gRNAs with second strand primers 100% complementary to a template region and positioned at the 3' end may be transcribed with a poly-U sequence immediately 3' of the second strand primer, inhibiting priming function. To solve this problem, gRNAs with cleavable RNA sequences positioned 3' of the second strand primer were developed to prevent formation of a

poly-U sequence at the 3' end of the second strand primer. gRNAs with either an HDV self-cleaving ribozyme or a tRNA positioned 3' of the second strand primer were tested.

[0318] The HDV self-cleaving ribozyme autocatalytically cleaved itself from the 3' end of the gRNA, leaving the second strand primer without a poly-U sequence. The HDV ribozyme left a 2'3' cyclic phosphate at the 3' end of the second strand primer which inhibited primer extension of the second strand primer. Endogenous polynucleotide kinase converted the 2'3' cyclic phosphate to a 3' OH, to enable primer extension. The tRNA was cleaved from the 3' end of the second strand primer by endogenous RNase P, leaving the second strand primer without a poly-U sequence and a 3' OH capable of primer extension.

[0319] **FIG. 18** shows the editing efficiency of a leucine zipper Cas9-RT construct in the presence of a gRNA comprising either an HDV ribozyme (left bar) or a tRNA (right bar) at the 3' end of the gRNA, immediately 3' of the second strand primer. The leucine zipper Cas9-RT construct was expressed as nCas9-LZ1 (SEQ ID NO: 1) and LZ2-mlvRT5M (SEQ ID NO: 2) and linked through a leucine zipper. The tRNA had a sequence corresponding to SEQ ID NO: 94. Editing was performed using gRNAs with an 85 nucleotide RT template, a 13 nucleotide primer binding site, a 19 nucleotide Velcro region, a 20 nucleotide second strand primer, and either an HDV ribozyme or a tRNA 3' of the second strand primer to edit a site such that an ATGG sequence, starting 2 nucleotides 3' of the nick, was mutated to CATA.

EXAMPLE 14

GPS-assisted reachover gRNAs (GARGs) and GPS-recruiting guides (GRGs)

[0320] GPS-assisted reachover gRNAs (GARGs) may improve a gene editing efficiency, for example in a case where a single guide nucleic acid system would otherwise generate a flap containing a desired edit that does not sufficiently displace a genomic strand that doesn't include the edit. A GARGs may be useful for promoting hybridization of the extended flap into the genome, and may anchor the 3' end of the extended flap in the vicinity of the genomic strand it is intended to replace. **FIG. 33** depicts data from a dual guide system. Two separate GARGs were tested and successfully led to gene editing in target nucleic acids of mammalian cells. Thus, a method employing a dual guide system comprising a GARG and a GRG may lead to precise genome editing in cells such as mammalian cells.

[0321] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives

to the embodiments of the disclosure described herein may be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS**WHAT IS CLAIMED IS:**

1. A method of increasing gene editing efficiency in a cell having a low deoxynucleoside triphosphate (dNTP) concentration and comprising a DNA polymerase for the gene editing, the method comprising:
increasing the dNTP concentration in the cell, relative to a baseline dNTP concentration.
2. The method of claim 1, wherein increasing the dNTP concentration in the cell comprises inhibiting a deoxynucleotide triphosphate triphosphohydrolase in the cell.
3. The method of claim 2, wherein the deoxynucleotide triphosphate triphosphohydrolase comprises SAM domain and HD domain-containing protein 1 (SAMHD1).
4. The method of claim 3, wherein inhibiting SAMHD1 comprises contacting the SAMHD1 with a Vpx protein, or expressing the Vpx protein in the cell.
5. The method of claim 3, wherein inhibiting SAMHD1 comprises contacting the SAMHD1 with a BGLF4 protein, or expressing the BGLF4 protein in the cell.
6. The method of claim 3, wherein inhibiting SAMHD1 comprises contacting an mRNA encoding the SAMHD1 with a microRNA or siRNA that hybridizes to the mRNA, or expressing the microRNA or siRNA in the cell.
7. The method of claim 3, wherein inhibiting SAMHD1 comprises contacting the SAMHD1 with a small molecule SAMHD1 inhibitor.
8. The method of claim 1, wherein increasing the dNTP concentration in the cell comprises administering nucleosides or nucleotides to the cell, wherein the nucleosides or nucleotides optionally comprise deoxynucleosides (dNs), deoxynucleoside monophosphates (dNMPs), or nucleoside triphosphates (NTPs).
9. The method of claim 8, wherein administering nucleosides or nucleotides to the cell comprises administering the nucleosides or nucleotides to a subject comprising the cell.
10. The method of claim 9, wherein the administration is oral or by injection.
11. The method of claim 1, wherein increasing the dNTP concentration in the cell comprises delivering a dNTP synthetic enzyme to the cell.
12. The method of claim 11, wherein the dNTP synthetic enzyme comprises a kinase.
13. The method of claim 12, wherein the kinase comprises a nucleoside kinase, deoxynucleoside kinase, deoxynucleoside monophosphate kinase, or deoxynucleotide diphosphate kinase.
14. The method of claim 1, wherein the DNA polymerase comprises a reverse transcriptase.

15. The method of claim 1, wherein the cell further comprises a Cas9 programmable nuclease, a guide nucleic acid, or a combination thereof.
16. The method of claim 1, wherein the low dNTP concentration comprises a dNTP concentration found in a nondividing cell.
17. The method of claim 1, wherein the low dNTP concentration is less than a dNTP concentration found in an activated peripheral blood mononuclear cell.
18. The method of claim 1, wherein the low dNTP concentration comprises a dNTP concentration below 1 micromolar.
19. The method of claim 1, wherein the increasing the dNTP concentration comprises increasing the dNTP concentration by at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, or more, relative to the baseline dNTP measurement.
20. The method of any one of claims 1-19, wherein the dNTP concentration comprises a deoxyadenosine triphosphate (dATP) concentration, a deoxycytidine triphosphate (dCTP) concentration, a deoxyguanosine triphosphate (dGTP) concentration, or a deoxythymidine triphosphate (dTTP) concentration, or any combination thereof.
21. A composition comprising a Cas nickase and a reverse transcriptase, wherein at least part of the Cas nickase and the reverse transcriptase are included in separate polypeptide chains, and wherein the Cas nickase and the reverse transcriptase form a Cas-reverse transcriptase heterodimer.
22. The composition of claim 21, wherein the Cas-reverse transcriptase heterodimer comprises a first heterodimer domain fused to the Cas nickase and a second heterodimer domain fused to the reverse transcriptase, wherein the first heterodimer domain binds the second heterodimer domain to form the Cas-reverse transcriptase heterodimer.
23. The composition of claim 22, wherein the first heterodimer domain is a leucine zipper and the second heterodimer domain is a leucine zipper.
24. The composition of any one of claims 21-23, wherein the reverse transcriptase comprises a sequence having at least 80% sequence identity to of any one of SEQ ID NO: 3 – SEQ ID NO: 22 or SEQ ID NO: 40 – SEQ ID NO: 80, or a fragment thereof.
25. The composition of any one of claims 21-24, wherein the reverse transcriptase comprises a domain from a non-long terminal repeat retrotransposable element fused to part of the Cas nickase.
26. The composition of any one of claims 21-24, wherein the reverse transcriptase comprises a sequence from a bacterial group II intron fused to part of the Cas nickase.

27. The composition of any one of claims 21-24, wherein the reverse transcriptase comprises a domain from a retroviral gag-pol polyprotein fused to part of the Cas nickase.
28. A composition comprising a Cas nickase, a reverse transcriptase, and a guide nucleic acid, wherein a first polypeptide comprises the Cas nickase and a second polypeptide comprises the reverse transcriptase and the guide nucleic acid binds to the Cas nickase and the reverse transcriptase.
29. The composition of any one of claims 21-28, wherein the reverse transcriptase comprises an mcp peptide.
30. The composition of any one of claims 21-29 wherein the reverse transcriptase comprises a loop region.
31. The composition of claim 30, wherein the loop region is a 2a loop or a 3a loop.
32. The composition of any one of claims 28-31, wherein the guide nucleic acid comprises a MS2 hairpin.
33. A composition comprising a reverse transcriptase with a sequence having at least 80% sequence identity to of any one of SEQ ID NO: 3 – SEQ ID NO: 22 or SEQ ID NO: 40 – SEQ ID NO: 80, or a fragment thereof fused to a Cas nickase.
34. A composition comprising a reverse transcriptase comprising a domain from a non-long terminal repeat retrotransposable element fused to a Cas nickase.
35. A composition comprising a reverse transcriptase comprising a sequence from a bacterial group II intron fused to a Cas nickase.
36. A composition comprising a reverse transcriptase comprising a domain from a retroviral gag-pol polyprotein fused to a Cas nickase.
37. A composition comprising a Cas nickase and a reverse transcriptase, wherein the Cas nickase and the reverse transcriptase comprise separate polypeptide chains, and wherein the Cas nickase and reverse transcriptase are not engineered to heterodimerize.
38. The composition of any one of claims 21-37, comprising a guide nucleic acid that forms a complex with the Cas nickase, wherein, upon complex formation, the Cas nickase is capable of introducing a single-strand break at a target site in a target nucleic acid.
39. The composition of any one of claims 21-38, wherein the target nucleic acid comprises a CFTR nucleic acid, a USH2A nucleic acid, an ABCA4 nucleic acid, an ATP7B nucleic acid, or an HTT nucleic acid.
40. The composition of any one of claims 21-39, comprising a nuclear localization signal fused to the Cas nickase or the reverse transcriptase.

41. The composition of any one of claims 21-40, wherein the reverse transcriptase is a truncated reverse transcriptase.
42. The composition of any one of claims 21-41, wherein the reverse transcriptase has an increased processivity as compared to a native reverse transcriptase.
43. The composition of any one of claims 21-42, wherein the reverse transcriptase has increased processivity compared to mlvRT.
44. The composition of any one of claims 21-43, wherein the reverse transcriptase edits a longer window length in a target sequence compared to mlvRT.
45. The composition of any one claims 21-44, wherein the reverse transcriptase has decreased immunogenicity compared to mlvRT.
46. The composition of any one claims 21-45, wherein the reverse transcriptase has improved delivery to a cell compared to mlvRT.
47. The composition of any one of claims 21-46, wherein the reverse transcriptase polymerizes 20 or more, 40 or more, 45 or more, 50 or more, 60 or more, 81 or more, 100 or more, 500 or more, or 1000 or more nucleotides in a single binding event.
48. A guide nucleic acid comprising:
a spacer reverse complementary to a first region of a target nucleic acid,
a scaffold configured to bind to a Cas nickase,
a reverse transcriptase template encoding a sequence to be inserted into the target nucleic acid, and
a first strand primer binding site reverse complementary to a second region of the target nucleic acid.
49. The guide nucleic acid of claim 48, further comprising a second strand primer comprising a sequence of a region of the reverse transcriptase template.
50. The guide nucleic acid of claim 48 or claim 49, wherein the first region of the target nucleic acid is on a first strand of the target nucleic acid and the second region of the target nucleic acid is on a second strand of the target nucleic acid.
51. The guide nucleic acid of any one of claims 48-50, wherein all or part of the first region of the target nucleic acid is reverse complementary to all or part of the second region of the target nucleic acid.
52. The guide nucleic acid of any one of claims 48-51, further comprising a cleavable sequence at the 3' end of the guide nucleic acid.
53. The guide nucleic acid of claim 52, wherein the cleavable sequence is a ribozyme cleavable sequence.

54. The guide nucleic acid of claim 52, wherein the cleavable sequence is a tRNA cleavable sequence.

55. The guide nucleic acid of any one of claims 48-54, wherein the first strand primer binding site is configured to hybridize to the second region of the target nucleic acid, and wherein the reverse transcriptase template is configured to serve as a template for reverse transcription from a 3' end of the second region of the target nucleic acid.

56. The guide nucleic acid of any one of claims 48-55, wherein the second strand primer is configured to serve as a primer for transcription from a template reverse complementary to the reverse transcriptase template.

57. The guide nucleic acid of any one of claims 48-56, wherein a first synthesized strand serves as a template for synthesis of a second strand from the second strand primer.

58. The guide nucleic acid of any one of claims 48-57, further comprising a Velcro region that hybridizes to a Velcro binding site.

59. The guide nucleic acid of claim 58, wherein the Velcro binding site is 100% reverse complementary to the Velcro region; wherein the Velcro binding site is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% reverse complementary to the Velcro region; and/or wherein the Velcro binding site is no greater than 55%, no greater than 60%, no greater than 65%, no greater than 70%, no greater than 75%, no greater than 80%, no greater than 85%, no greater than 90%, no greater than 91%, no greater than 92%, no greater than 93%, no greater than 94%, no greater than 95%, no greater than 96%, no greater than 97%, no greater than 98%, no greater than 99% reverse complementary to the Velcro region.

60. The guide nucleic acid of claim 58 or 59, wherein the reverse transcriptase template region comprises the Velcro binding site.

61. The guide nucleic acid of claim 58 or 59, wherein the Velcro binding site is 3' of the first strand primer binding site.

62. The guide nucleic acid of any one of claims 48-61, wherein the Velcro region is 3' of the reverse transcriptase template.

63. The guide nucleic acid of any one of claims 48-62, wherein the Velcro region is 5' of the scaffold.

64. The guide nucleic acid of any one of claims 48-63, wherein the target nucleic acid comprises a CFTR nucleic acid, a USH2A nucleic acid, an ABCA4 nucleic acid, an ATP7B nucleic acid, or an HTT nucleic acid.

65. The guide nucleic acid of any one of claims 48-64, wherein the spacer comprises a nucleic acid sequence at least 85% identical to any one of SEQ ID NOs: 96-119.
66. A composition comprising a first guide nucleic acid comprising the guide of any one of claims 28-32 or 37-65 and a second guide nucleic acid.
67. The composition of claim 66, wherein the second guide nucleic acid comprises the guide nucleic acid any one of claims 28, 32 or 37, or 48-65.
68. The composition of claim 67, wherein the reverse transcriptase template of the second guide nucleic acid is complementary (or at least partly complementary) to at least part of the reverse transcriptase template of the first guide nucleic acid.
69. The composition of any one of claims 66-68, wherein the first guide nucleic acid binds to a first Cas nickase, and the second guide nucleic acid binds to a second Cas nickase.
70. The composition of any one of claims 66-68, wherein a first spacer of the first guide nucleic acid binds a first Cas nickase, a second spacer of the second guide nucleic acid binds a second Cas nickase, a first scaffold of the first guide nucleic acid binds the second Cas nickase, and a second scaffold of the second guide nucleic acid binds the first Cas nickase.
71. The composition of any one of any one of claims 66-68 or 70, wherein the first guide nucleic acid comprises a first linker and the second guide nucleic acid comprises a second linker, wherein the first linker hybridizes to the second linker.
72. A method of increasing genome editing efficiency comprising delivering an Orf1p to a cell expressing the composition of any one of claims 21-47 or 66-71 or the guide nucleic acid of any one of claims 38-45.
73. One or more nucleic acids encoding the composition of any one of claims 21-47 or 66-71, or comprising the guide nucleic acid of any one of claims 48-65.
74. A viral vector comprising the nucleic acid of claim 73.
75. A cell comprising the composition of any one of claims 21-47 or 66-71, the guide nucleic acid of any one of claims 48-65, the nucleic acid of claim 73, or the viral vector of claim 74.
76. The method of claim 72 or the cell of claim 75, wherein the cell is a prokaryotic cell.
77. The method of claim 72 or the cell of claim 75, wherein the cell is a eukaryotic cell.
78. A method of increasing genome editing efficiency comprising expressing a Vpx protein in a cell.

79. The method of claim 78, wherein the cell expresses the composition of any one of claims 21-47 or 66-71 or the guide nucleic acid of any one of claims 48-65.

80. A method of increasing genome editing efficiency by increasing the dNTP concentration in a cell, for example a method of increasing genome editing efficiency comprising inhibiting SAMHD1 in a cell.

81. The method of claim 80, wherein the cell expresses a Cas9 programmable nuclease, a reverse transcriptase, and a guide nucleic acid.

82. The method of claim 80 or 81, wherein inhibiting SAMHD1 comprises expressing a Vpx protein in the cell.

83. The method of claim 80 or 81, wherein inhibiting SAMHD1 comprises expressing a microRNA against SAMHD1 in the cell, or comprises treating the cell with a small molecule SAMHD1 inhibitor.

84. A composition comprising a Cas9 programmable nuclease comprising one or more point mutations or insertion mutations that enable or improve intein catalysis.

85. The composition of claim 84, wherein the Cas9 programmable nuclease comprises a point mutation or insertion mutation located in a C-terminal half of the Cas9 programmable nuclease, or wherein the point mutation or insertion mutation is located anywhere after amino acid position 574 of the Cas9 programmable nuclease.

86. The composition of claim 85, wherein the point mutation comprises a cysteine point mutation, a serine point mutation, a threonine point mutation, or an alanine point mutation; or wherein the insertion mutation comprises a cysteine insertion mutation, a serine insertion mutation, a threonine insertion mutation, or an alanine insertion mutation.

87. The composition of claim 85, wherein the point mutation comprises a cysteine point mutation, or wherein the insertion mutation comprises a cysteine insertion mutation.

88. The composition of any one of claims 84-87, wherein the Cas9 programmable nuclease is a Cas9 nickase.

89. The composition of any one of claims 84-88, wherein the Cas9 programmable nuclease is an *S. Pyogenes* Cas9.

90. The composition of claim 89, wherein the point mutation is located at D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212 of the *S. Pyogenes* Cas9, or wherein the insertion mutation is located immediately upstream of D1079, D1125, D1130, G1133, A1140, I1168, S1173, D1180, G1186, L1203, or R1212 of the *S. Pyogenes* Cas9.

91. The composition of any one of claims 84-90, wherein the Cas9 programmable nuclease comprises a sequence of any one of SEQ ID NO: 85 – SEQ ID NO: 87 or SEQ ID NO: 90 – SEQ ID NO: 92.

92. The composition of any one of claims 84-91, wherein the Cas9 programmable nuclease is expressed as two or more segments.

93. The composition of claim 92, wherein a first segment of the two or more segments comprise an N-terminal portion of the Cas9 programmable nuclease and a first intein, and wherein a second segment of the two or more segments comprise a C-terminal portion of the Cas9 programmable nuclease and a second intein.

94. The composition of claim 93, wherein the cysteine point mutation is located at the N-terminus of the C-terminal portion of the Cas9 programmable nuclease.

95. The composition of claim 93 or 94, wherein the first intein is fused to the C-terminus of the N-terminal portion of the Cas9 programmable nuclease, and wherein the second intein is fused to the N-terminus of the C-terminal portion of the Cas9 programmable nuclease.

96. The composition of any one of claims 93-95, wherein the first segment comprises a sequence of SEQ ID NO: 90, and wherein the second segment comprises a sequence of SEQ ID NO: 91.

97. The composition of any one of claims 93-96, wherein the second segment of the two or more segments comprise a reverse transcriptase fused to the C-terminal portion of the Cas9 programmable nuclease.

98. The composition of claim 97, wherein the reverse transcriptase comprises an N-terminus fused to a C-terminus of the C-terminal portion of the Cas9 programmable nuclease.

99. The composition of claim 97 or 98, wherein the reverse transcriptase comprises an mlvRT, or a variant thereof.

100. A method of optimizing genome editing efficiency, comprising performing genome editing with a Moloney leukemia virus reverse transcriptase (mlvRT) that is modified to increase its catalytic efficiency in low dNTP concentrations, (e.g. modified to decrease its K_m for dNTPs).

101. A method of optimizing genome editing efficiency in a limiting dNTP condition, comprising performing genome editing with a Moloney leukemia virus reverse transcriptase (mlvRT), or a variant thereof, comprising a point mutation at position 221 or 223 of the reverse transcriptase.

102. The method of claim 100 or 101, wherein the mlvRT or variant thereof comprises a point mutation at position 221.

103. The method of claim 102, wherein the point mutation at position 221 comprises Q221R.
104. The method of claim 100 or 101, wherein the mlvRT or variant thereof comprises a point mutation at position 223.
105. The method of claim 104, wherein the point mutation at position 223 comprises V223A.
106. The method of claim 104, wherein the point mutation at position 223 comprises V223M.
107. The composition of any of claims 21-47 or 66-71, wherein the reverse transcriptase comprises a point mutation at position P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524.
108. The composition of any of claims 21-47 or 66-71, wherein the reverse transcriptase comprises a point mutation comprising P51L, S67R, Q84A, L139P, Q221R, V223A, V223M, T197A, D653N, T664N, L671P, L435G, H204R, or D524A.
109. The composition of any of claims 21-47 or 66-71, wherein the reverse transcriptase comprises a point mutation at amino acid position Q84, L139, Q221, V223, T664, or L671.
110. The composition of any of claims 21-47 or 66-71, wherein the reverse transcriptase comprises a point mutation comprising S67R, Q84A, L139P, Q221R, V223A, V223M, T664N, L671P, or D524A.
111. The composition of any of claims 21-47, wherein the Cas nickase and RT are encoded by polynucleotides.
112. AAVs comprising the polynucleotides of claim 111.
113. The AAVs of claim 112, wherein at least part of the Cas nickase and RT are encompassed by separate AAVs.
114. Adeno-associated viruses (AAVs) comprising: a first AAV comprising a first polynucleotide encoding at least part of a Cas nickase, and a second AAV comprising a second polynucleotide encoding a reverse transcriptase.
115. The AAVs of any one of claims 112-114, wherein the AAVs comprise AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV-DJ, AAV-DJ/8, AAV-Rh10, AAV-Rh74, AAV-retro, AAV-PHP.B, AAV8-PHP.eB, or AAV-PHP.S, or a combination of thereof.
116. The AAVs of claim 114 or 115, wherein the Cas nickase and the reverse transcriptase form a heterodimer with each other.

117. The AAVs of any one of claims 114-116, wherein the first or second polynucleotide further encodes a guide nucleic acid that binds to the Cas nickase and the reverse transcriptase to form a complex, and wherein the Cas nickase of the complex introduces a single-strand break at a target site in a target nucleic acid.

118. The AAVs of claim any one of claims 114-117, wherein the Cas nickase comprises a Cas9 nickase such as an *S. Pyogenes* Cas9 nickase, and the reverse transcriptase comprises an mlvRT, or a variant thereof, wherein the reverse transcriptase comprises a point mutation at P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524.

119. The AAVs of claim 118, wherein the point mutation comprises P51L, S67R, Q84A, L139P, Q221R, V223A, V223M, T197A, D653N, T664N, L671P, L435G, H204R, or D524A.

120. The AAVs of any one of claims 114-119, wherein the Cas9 nickase comprises an *S. Pyogenes* Cas9 nickase, and the reverse transcriptase comprises an mlvRT, or a variant thereof, wherein the reverse transcriptase comprises an insertion mutation immediately upstream of P51, S67, Q84, L139, Q221, V223, T197, D653, T664, L671, L435, H204, or D524.

121. A method of genome editing, comprising administering a composition comprising the first or second AAV of any one of claims 114-120 to a subject or cell.

122. A method of genome editing, comprising administering a composition comprising the AAVs of any one of claims 112-120 to a subject or cell.

123. The method of claim 121 or 122, further comprising measuring genome editing in the subject or cell.

124. A method of increasing gene editing efficiency in a cell having a low deoxynucleoside triphosphate (dNTP) concentration, comprising:

contacting the cell with a gene editing enzyme modified for efficient catalysis in the low dNTP concentration, or expressing the gene editing enzyme in the cell.

125. The method of claim 124, wherein the gene editing enzyme comprises a reverse transcriptase that is modified by introducing a point mutation at position Q84, L139, Q221, V223, T664, or L671.

126. A method for screening or identifying an improved reverse transcriptase (RT), comprising:

overexpressing SAMHD1, or expressing a mutant SAMHD1 that has been mutated to prevent phosphorylation of a residue of the mutant SAMHD1, in cells;

identifying an RT activity in the cells; and

based on the RT activity, identifying the RT as an improved RT.

127. A system comprising an RNA or polynucleotide comprising a spacer, a reverse transcriptase template comprising a desired edit, and a primer binding site, in which the primer binding site binds to a nucleic acid that does not comprise any part of the region of the nucleic acid targeted or bound by the spacer or the nucleic acid reverse complementary to the nucleic acid targeted or bound by the spacer.

128. A system comprising:

a first guide nucleic acid comprising:

a spacer reverse complementary to a first region of a target nucleic acid;

a scaffold configured to bind to a Cas nuclease;

a reverse transcriptase template encoding a sequence to be inserted into the target nucleic acid;

a first strand primer binding site that binds to a region of the target nucleic acid that does not comprise any part of the first region, and that does not comprise any part of a reverse complement of the first region; and

a GPS region that hybridizes to a GPS binding site on a second guide nucleic acid.

129. The system of claim 128, further comprising the second guide nucleic acid comprising the GPS binding site.

130. The system of claim 129, wherein the second guide nucleic acid comprises a second spacer reverse complementary to another region of the target nucleic acid.

131. The system of claim 129 or 130, wherein the second guide nucleic acid brings the primer binding site into close proximity with a genomic flap.

1/48

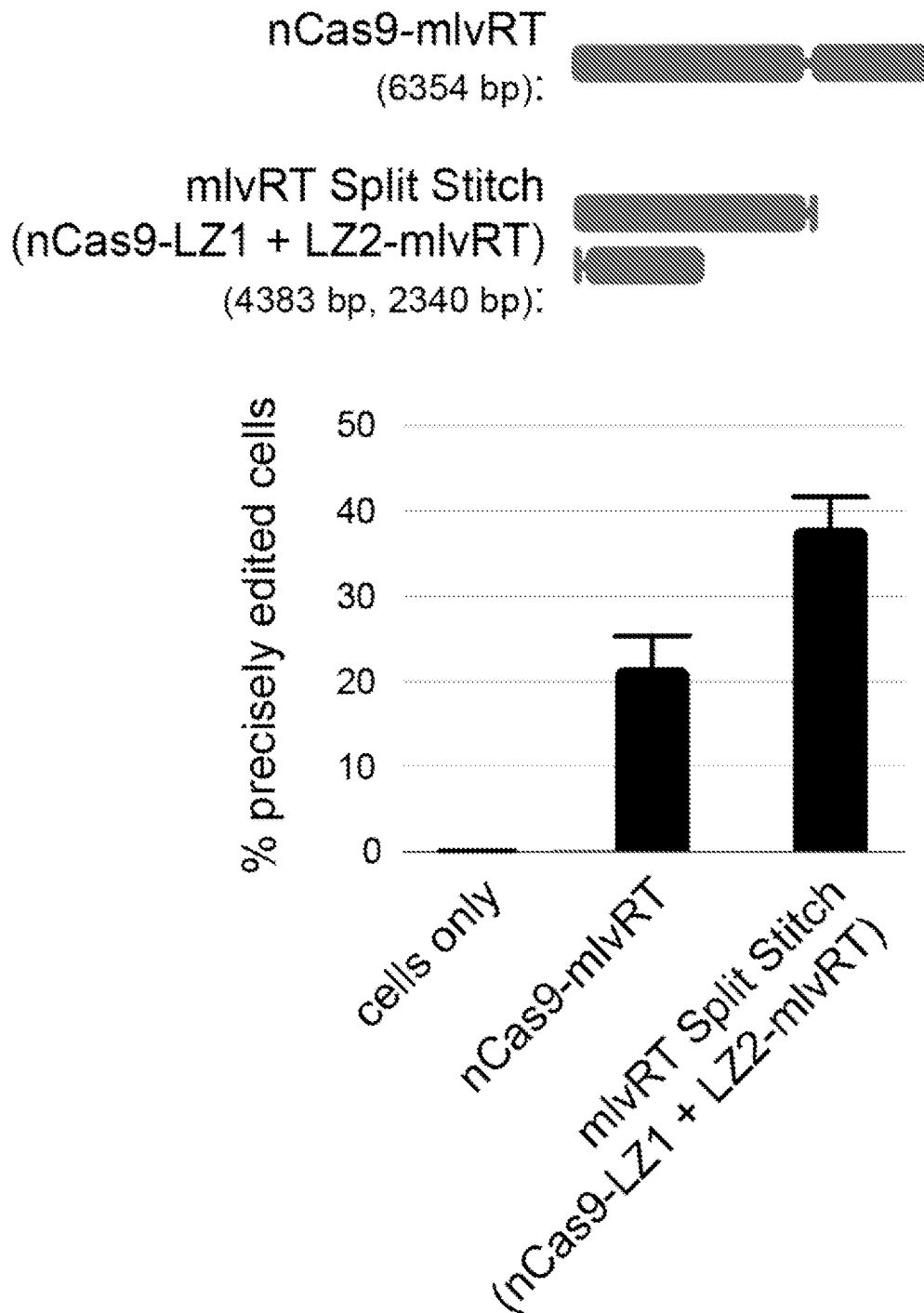


FIG. 1

2/48

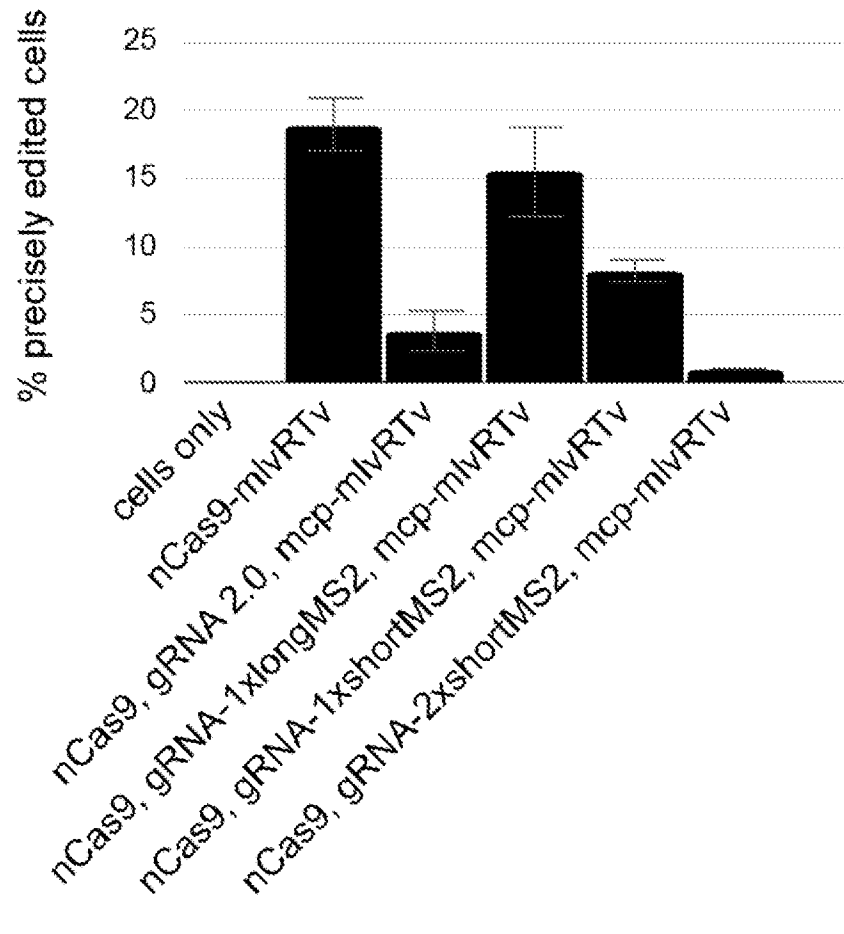


FIG. 2

3/48

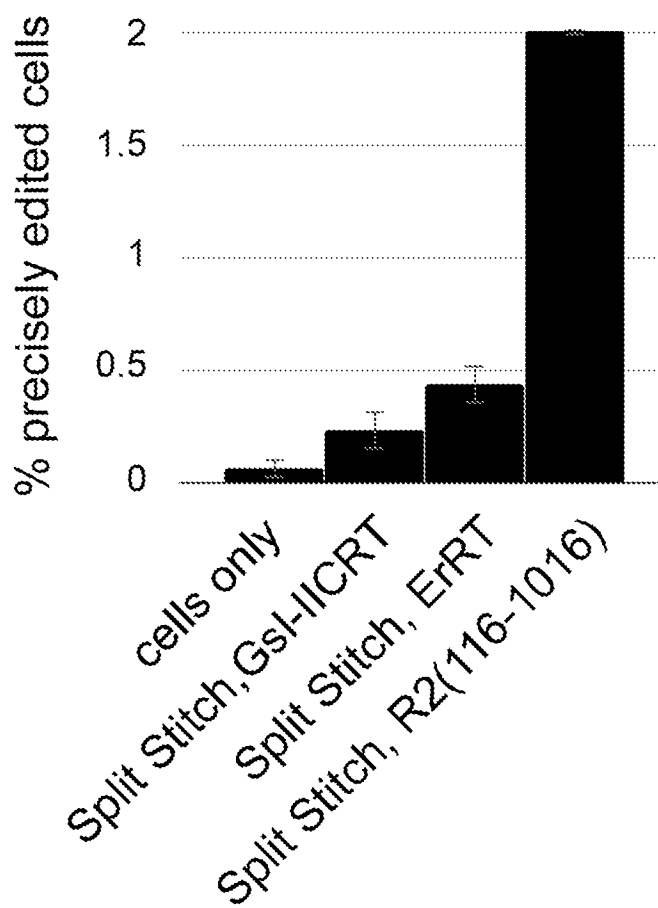
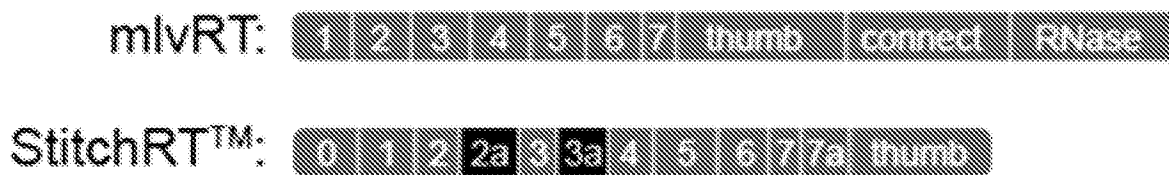


FIG. 3

4/48

FIG. 4A

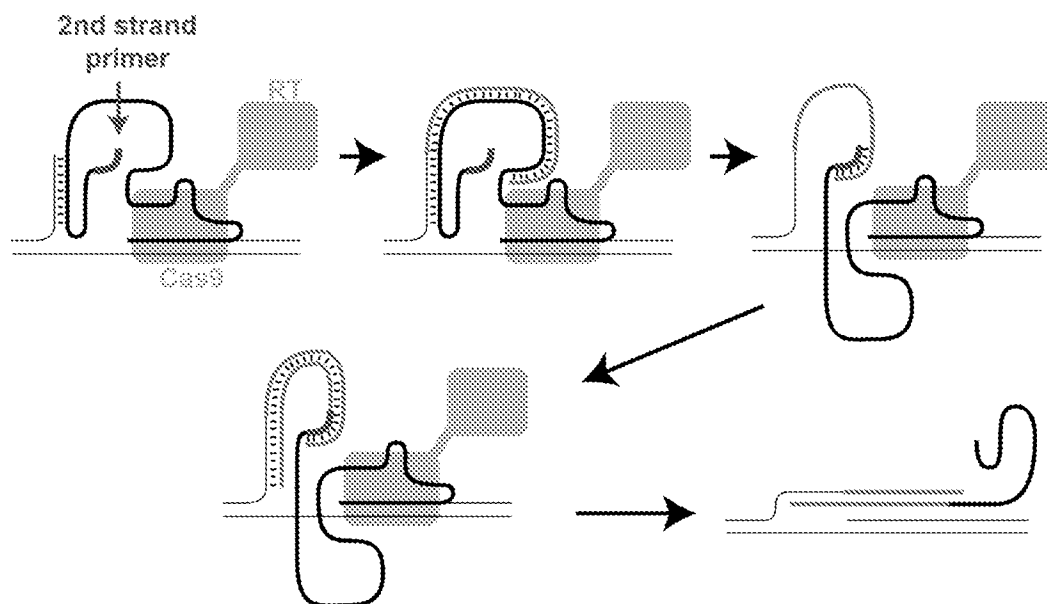


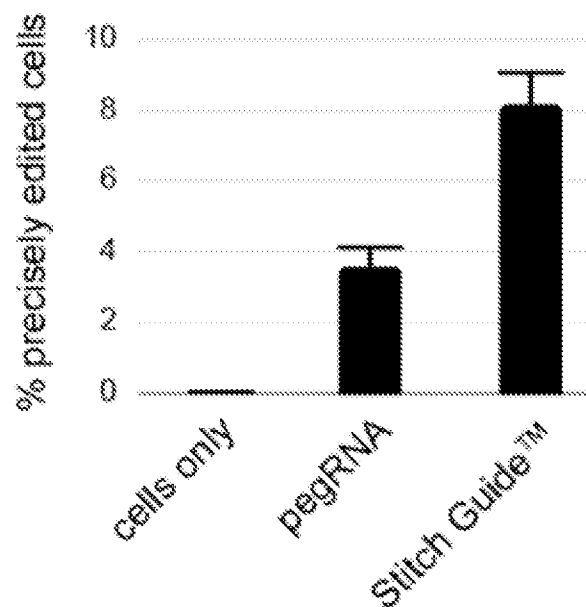
FIG. 4B

pegRNA:

spacer	scaffold	RT template	1 st strand PFS
--------	----------	-------------	----------------------------

Stitch Guide™:

spacer	scaffold	RT template	1 st strand PFS	2 nd strand primer
--------	----------	-------------	----------------------------	-------------------------------



5/48

FIG. 5A

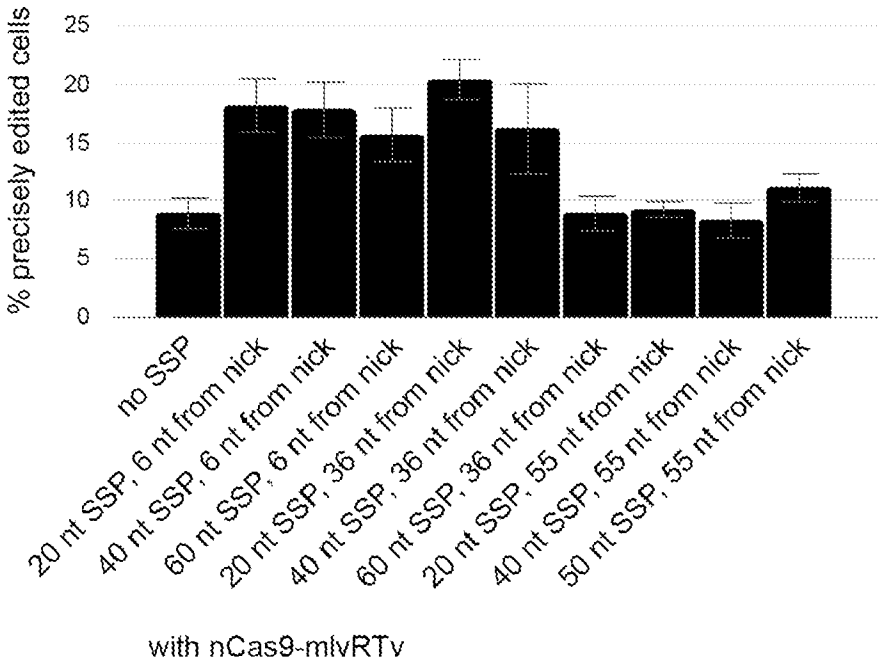
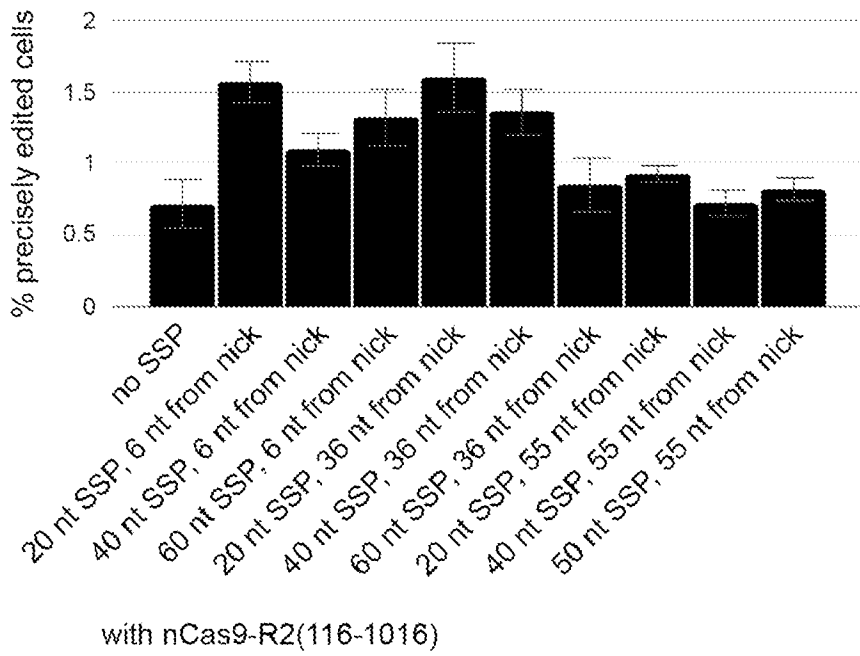


FIG. 5B



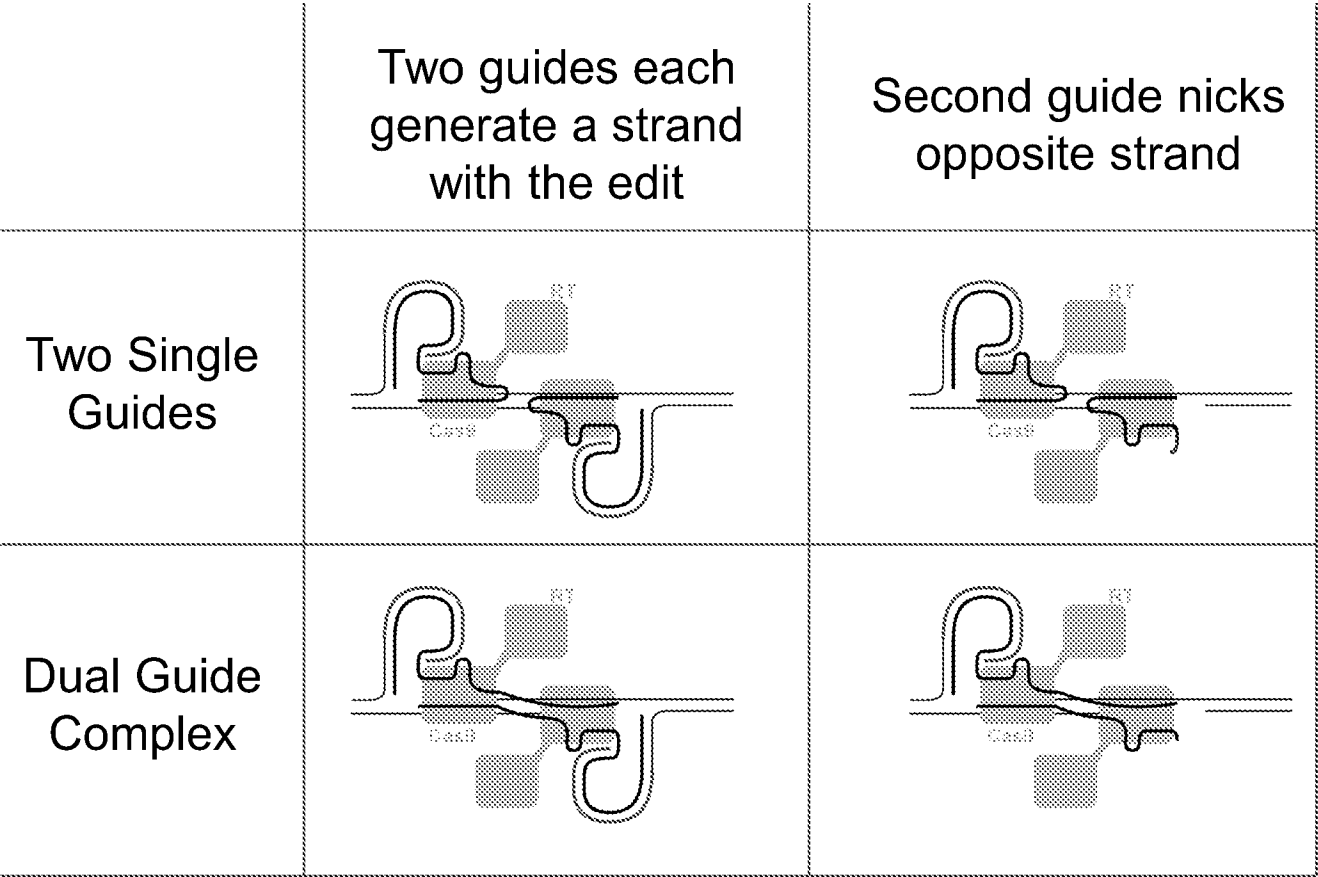


FIG. 6

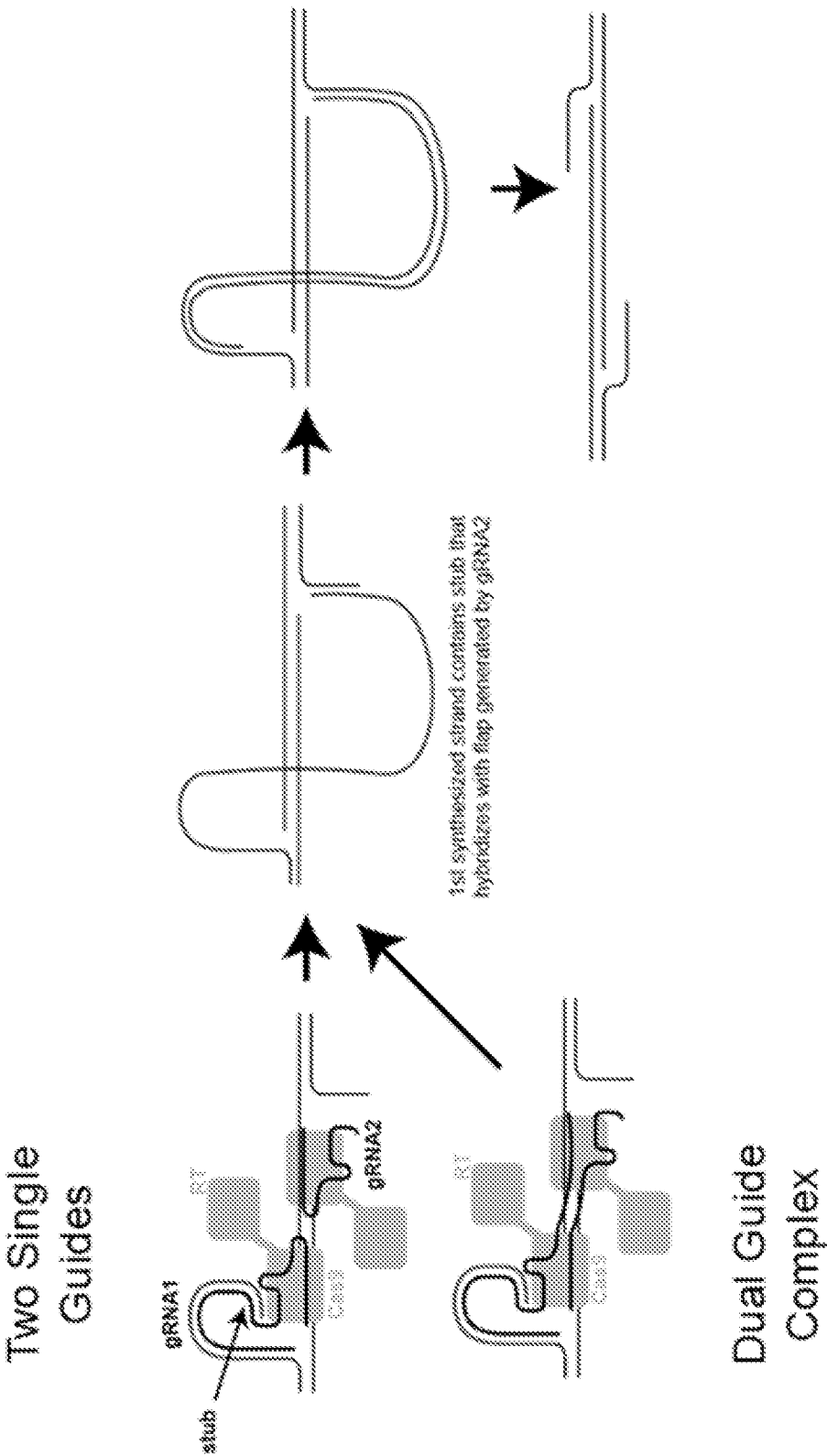


FIG. 7

FIG. 8A

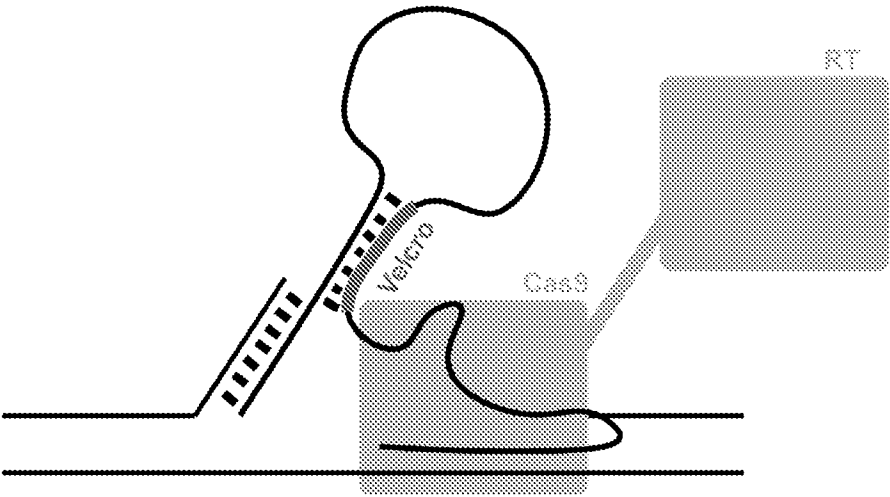
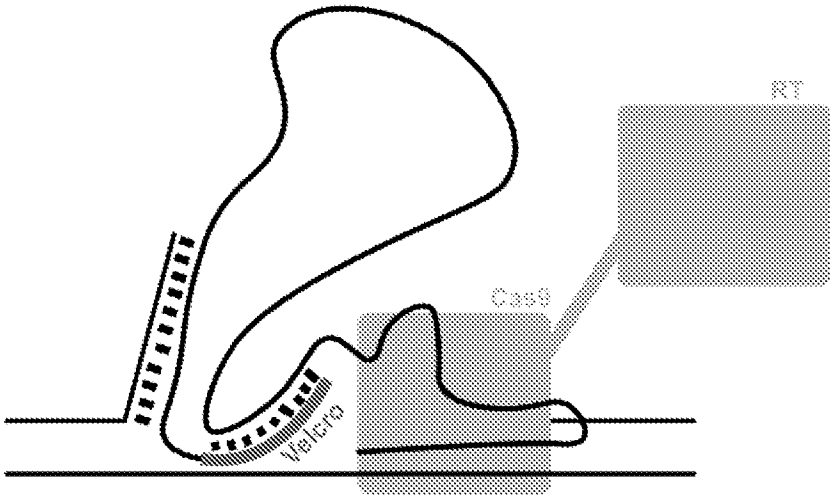


FIG. 8B



9/48

FIG. 9A

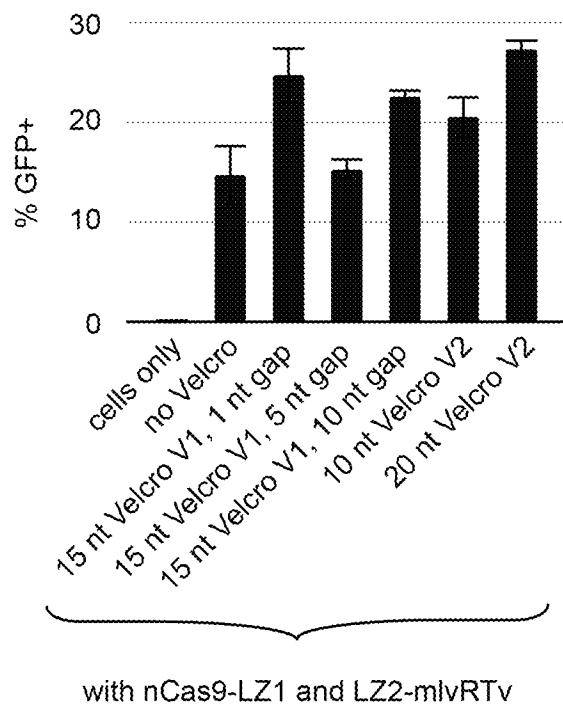
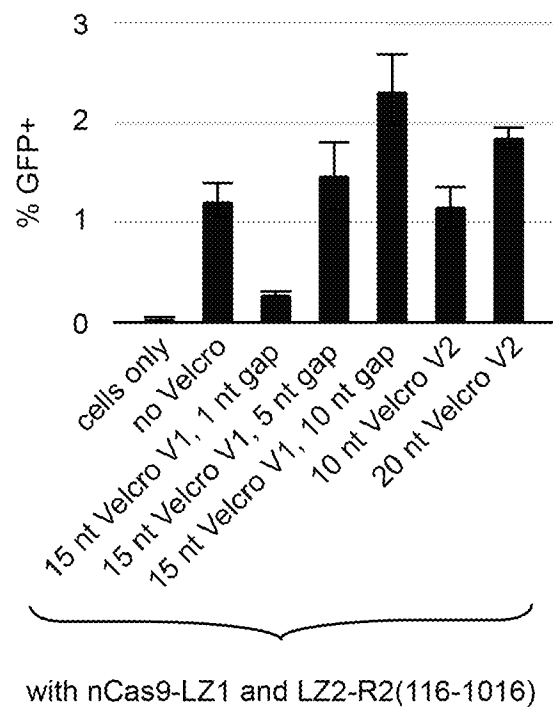


FIG. 9B



10/48

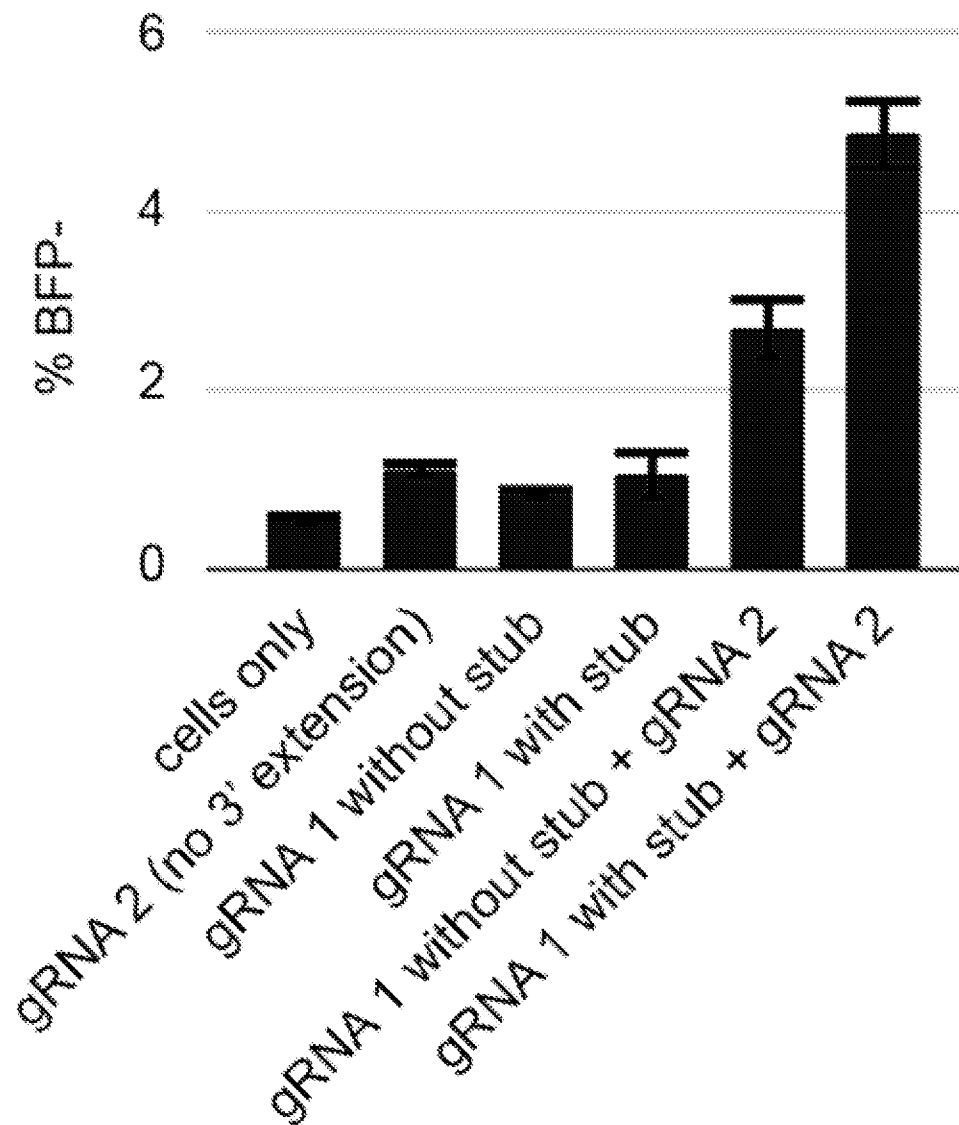


FIG. 10

11/48

FIG. 11A

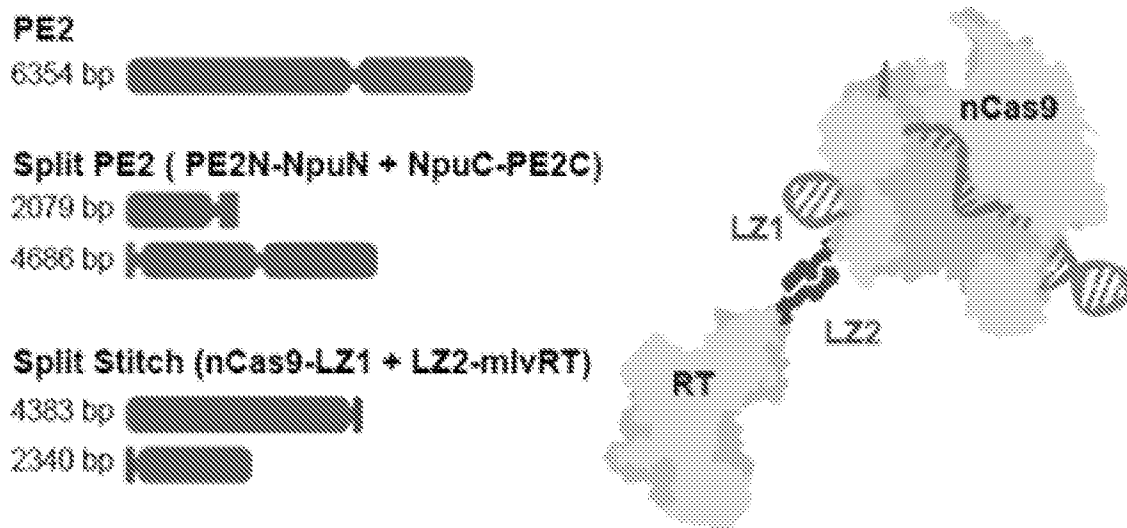


FIG. 11B

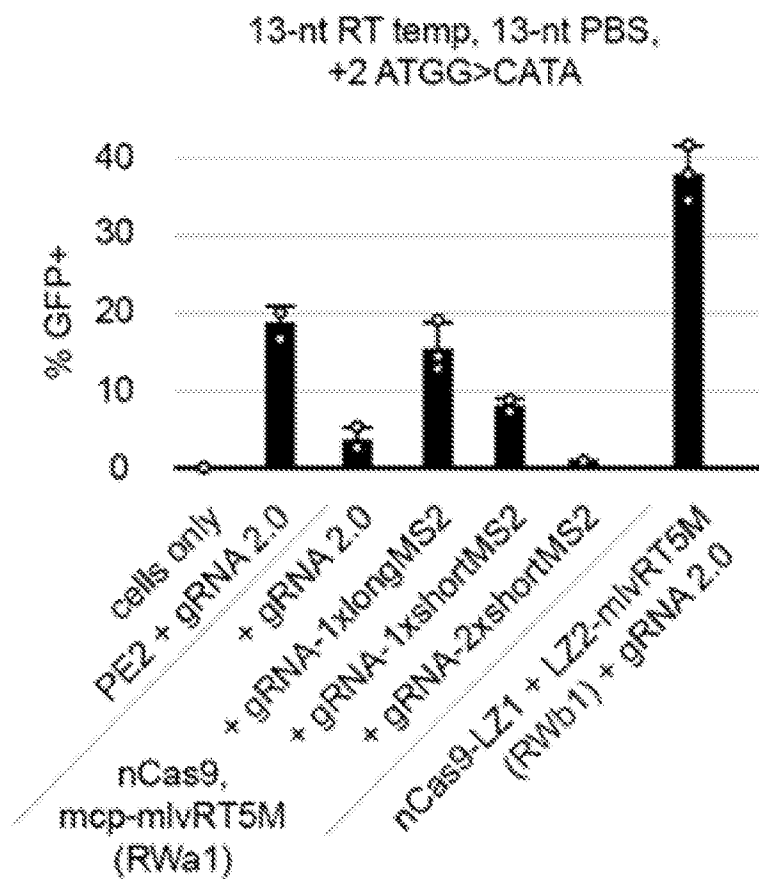


FIG. 12A

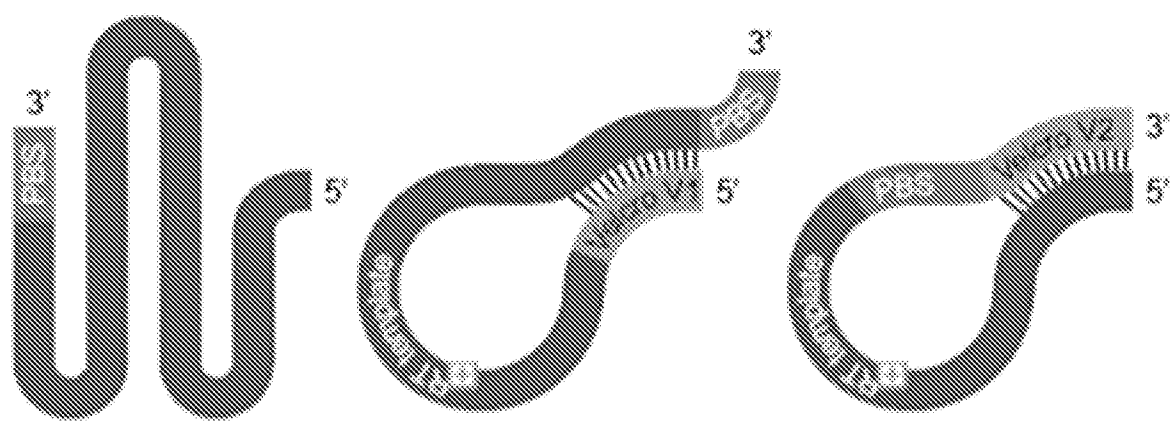


FIG. 12B

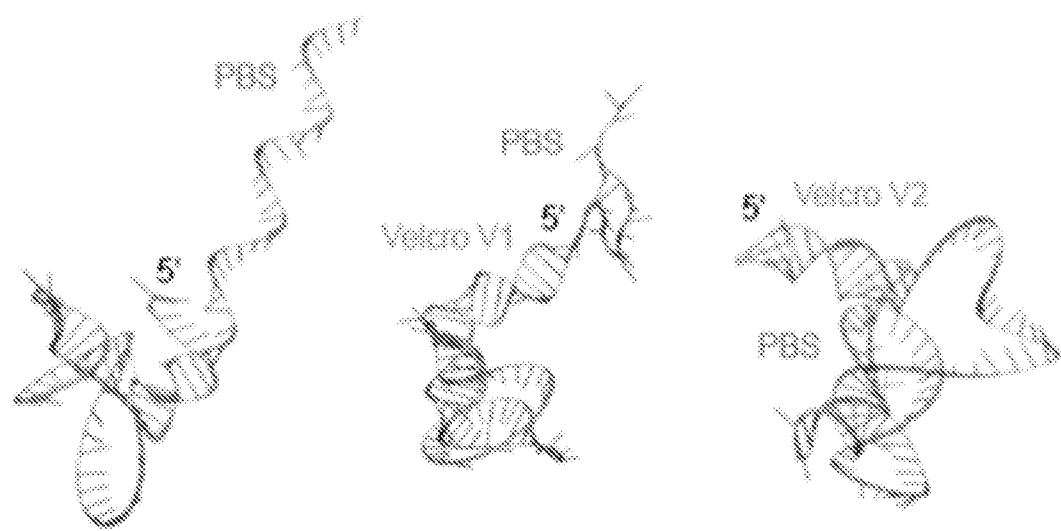


FIG. 12C

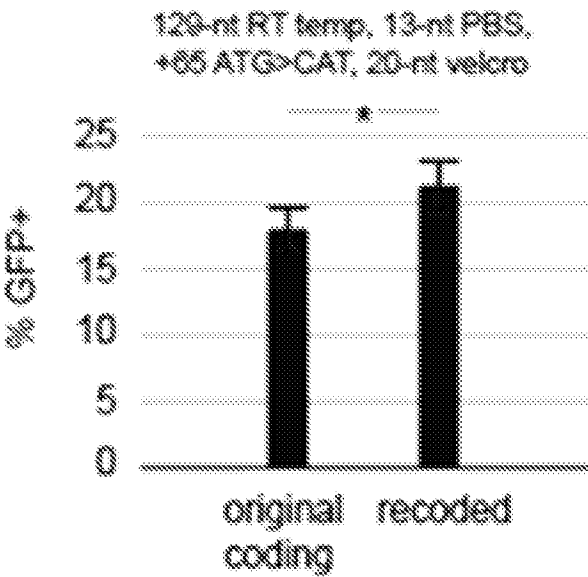


FIG. 12D

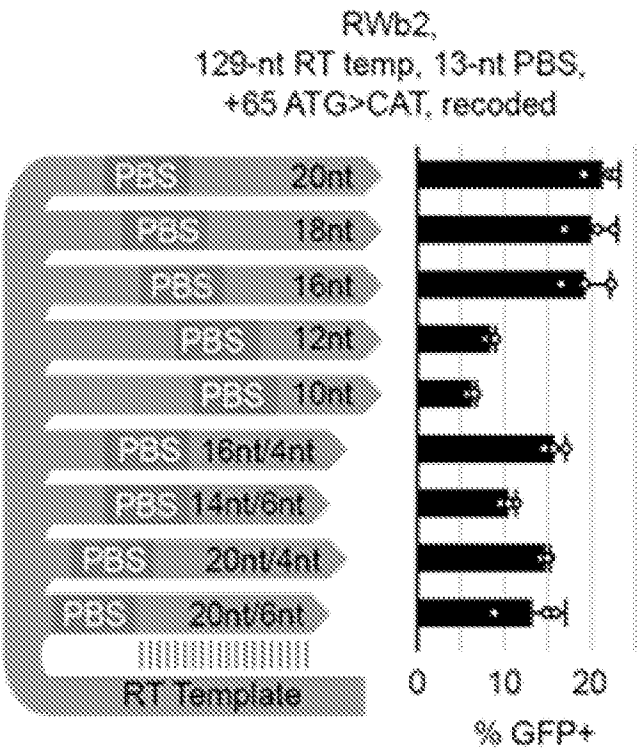


FIG. 13A

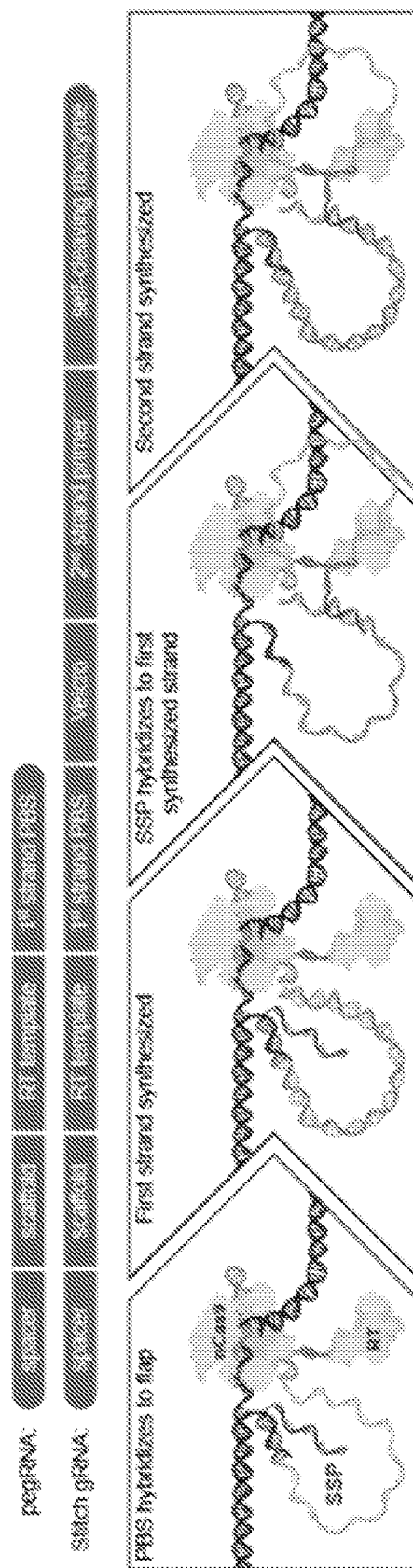


FIG. 13C

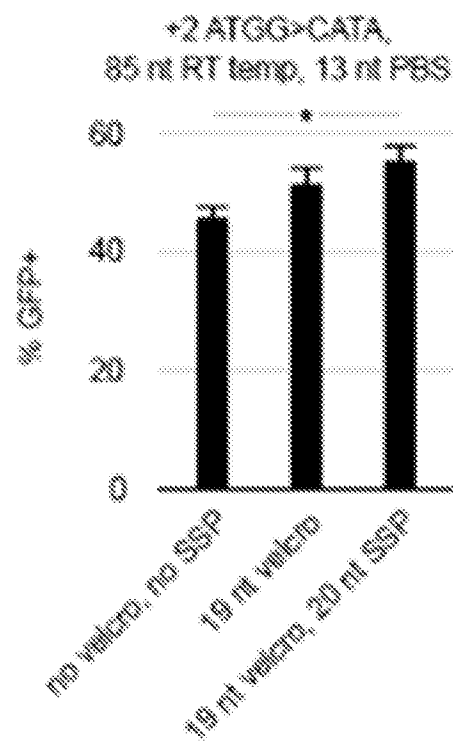


FIG. 13B

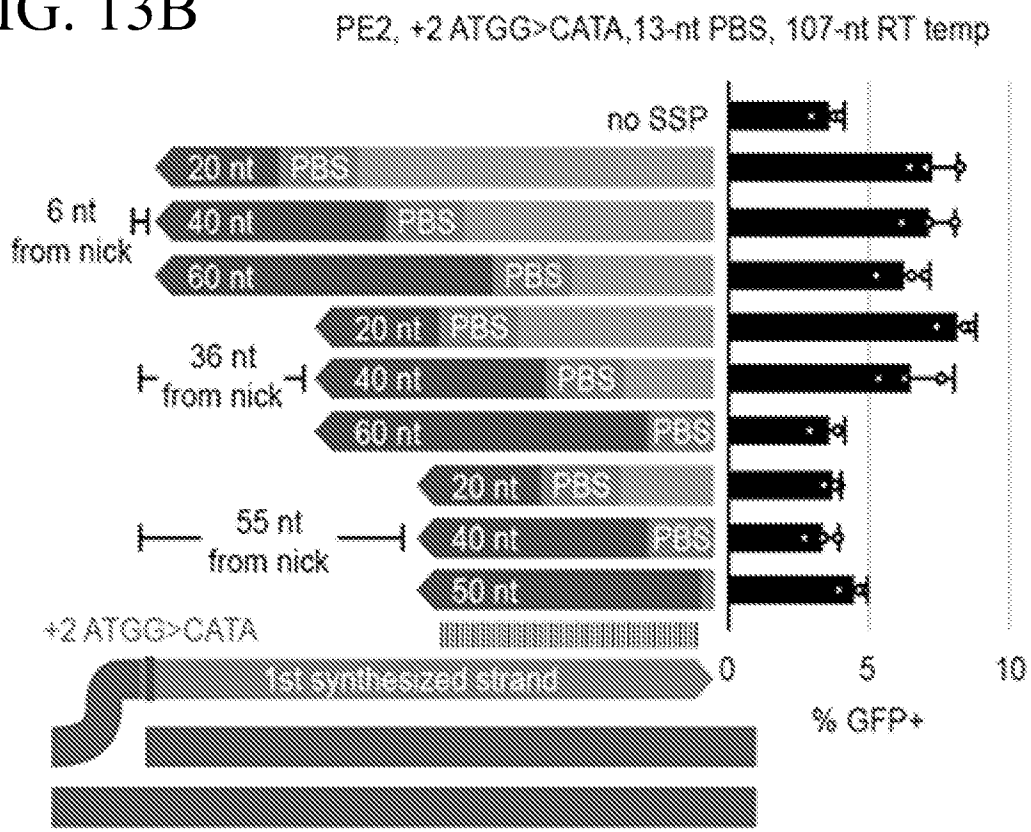
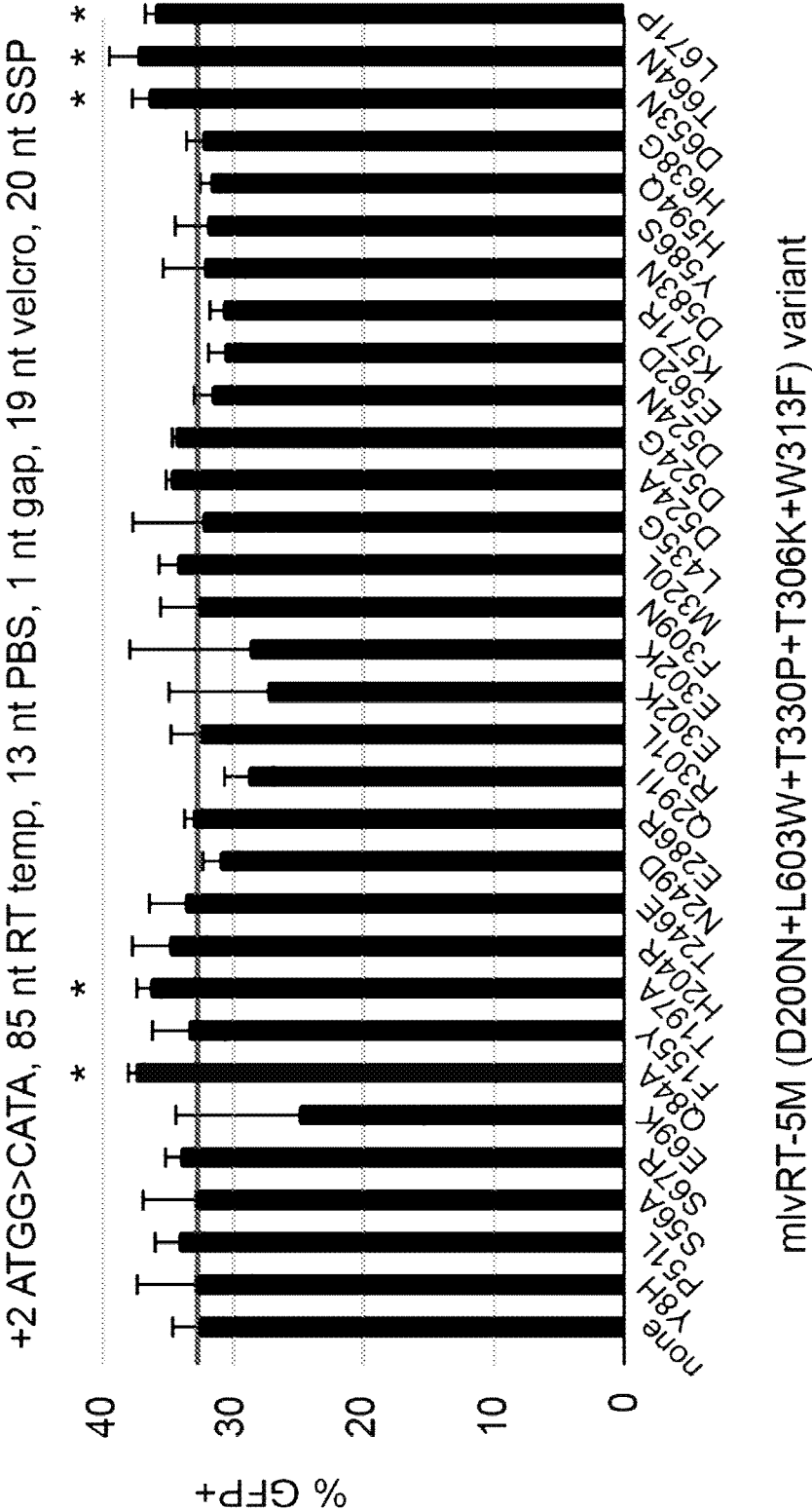


FIG. 14A



17/48

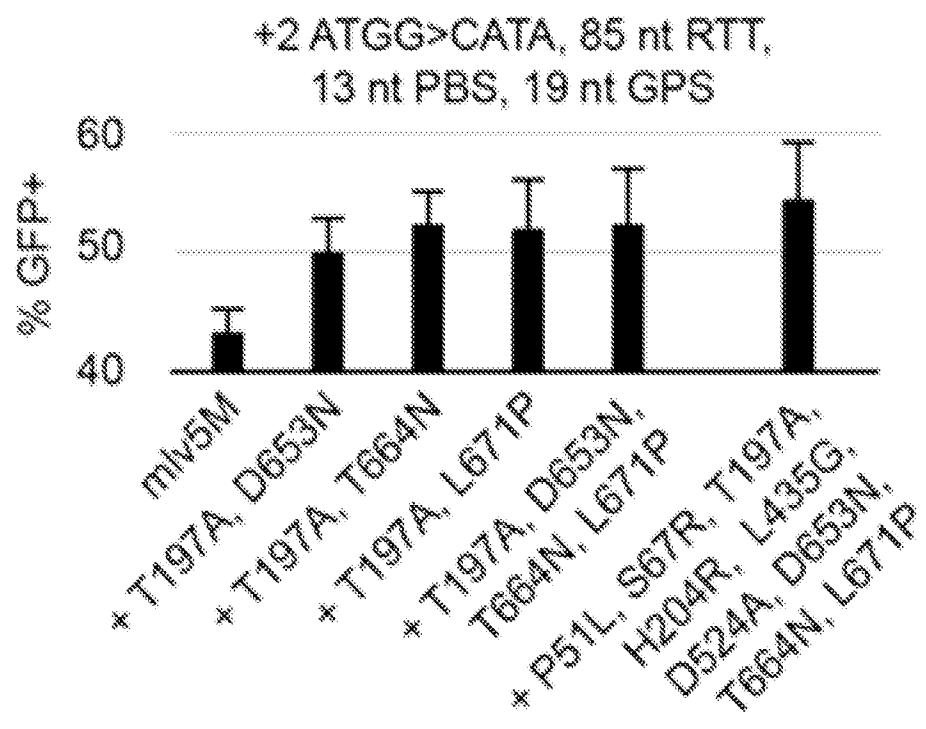


FIG. 14B

18/48

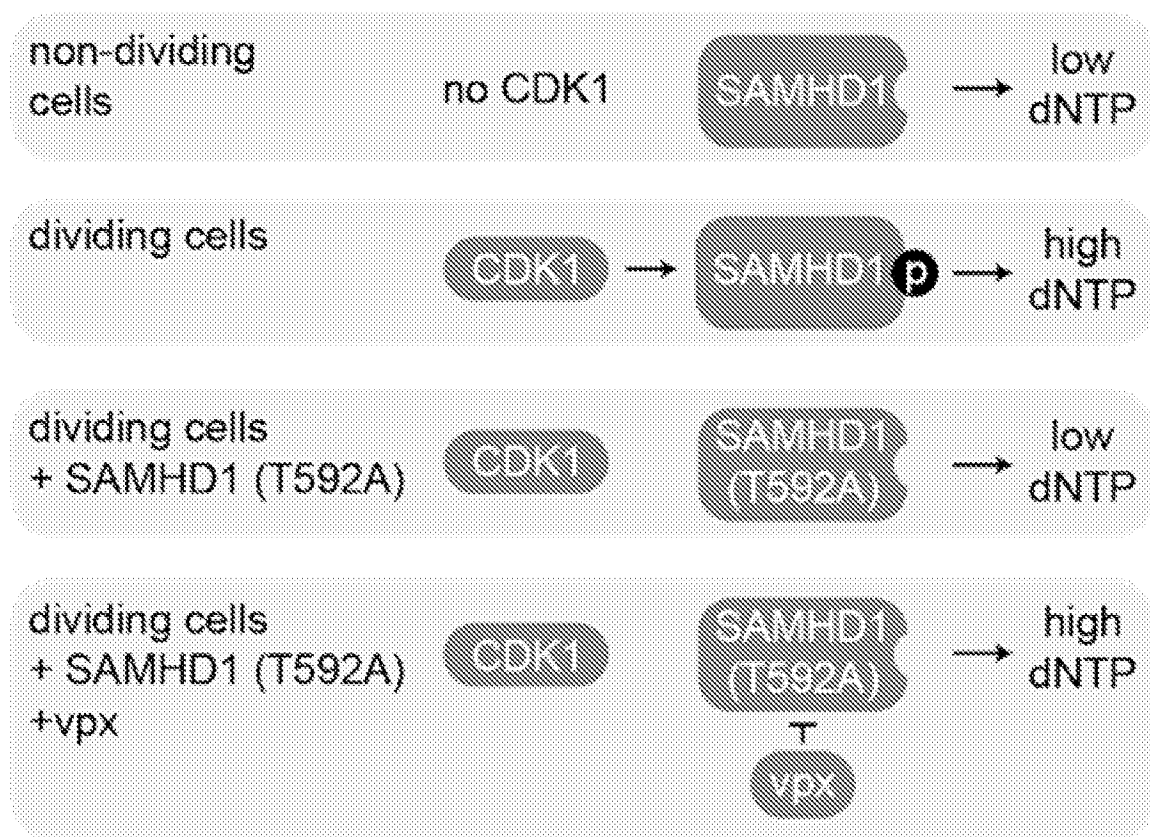


FIG. 15A

FIG. 15B

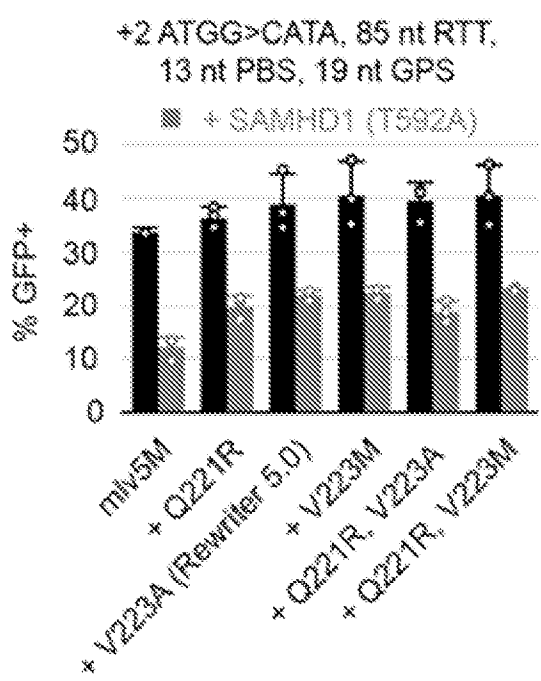


FIG. 15C

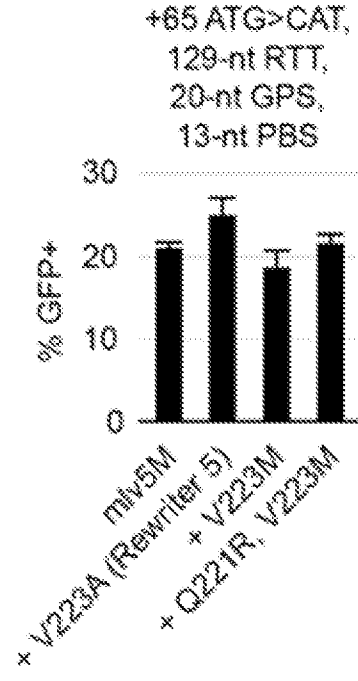


FIG. 15D

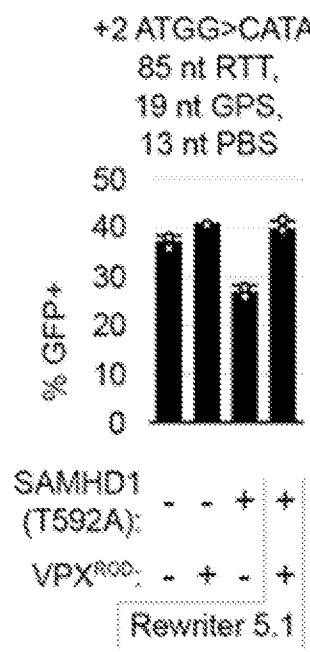


FIG. 15E

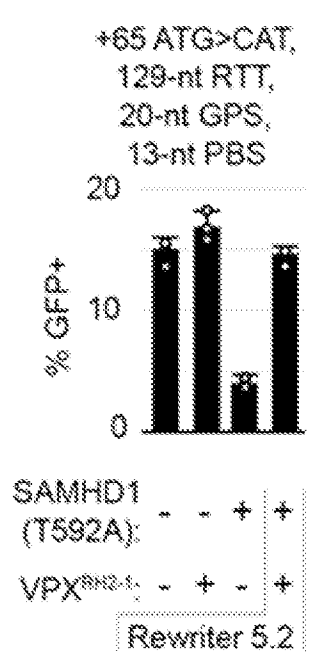
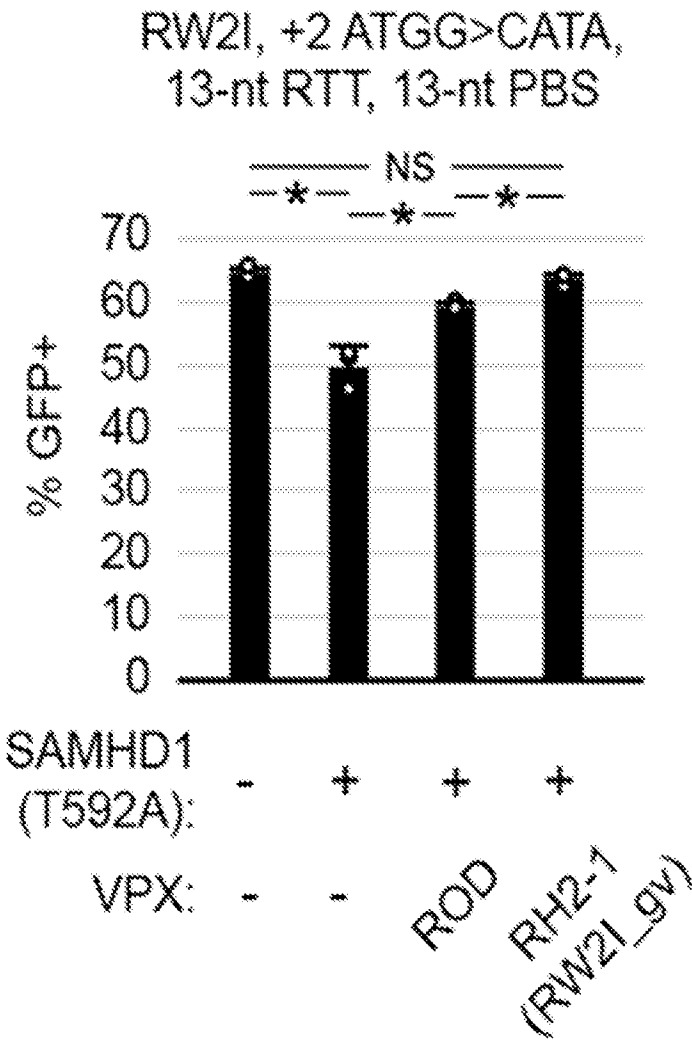


FIG. 15F



21/48

+2 ATGG>CATA,
85 nt RT temp,
13 nt PBS,
19 nt velcro, 20 nt SSP

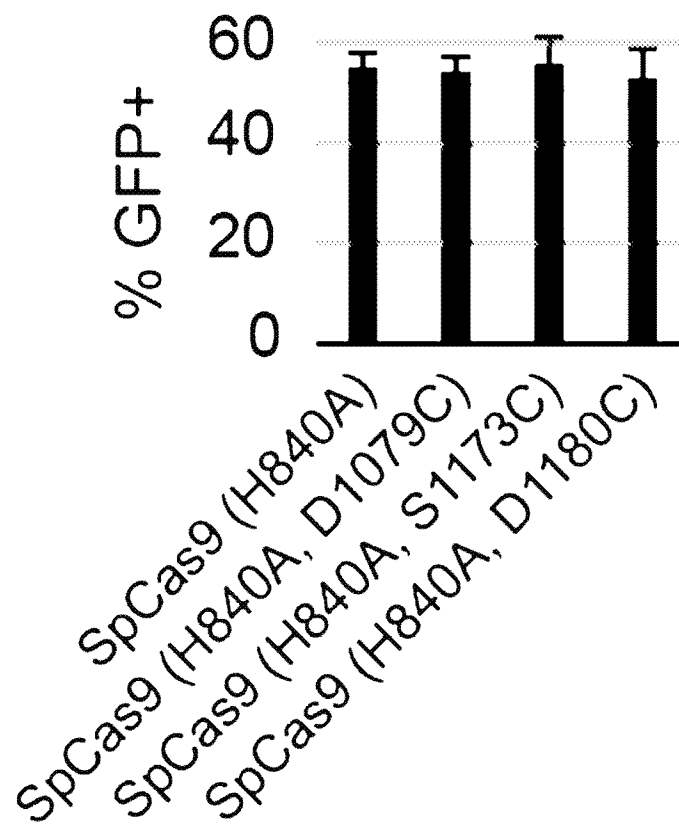


FIG. 16

22/48

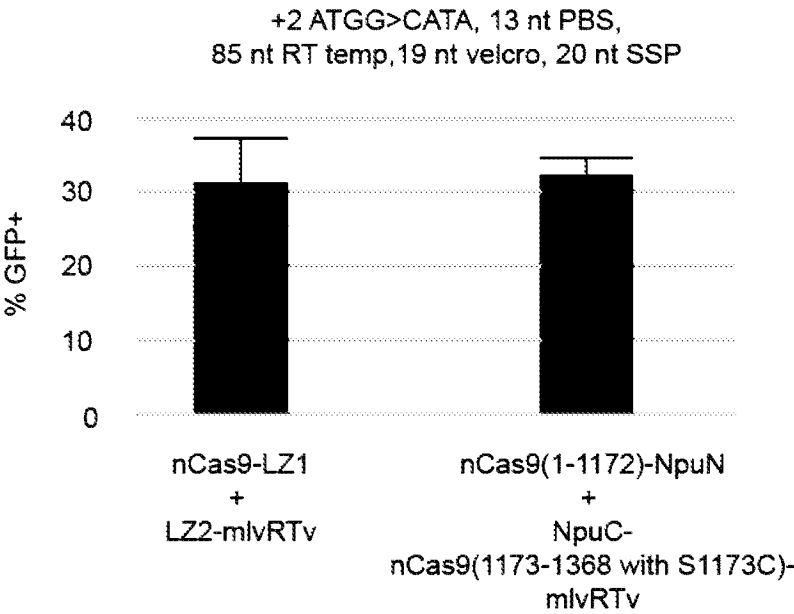


FIG. 17A

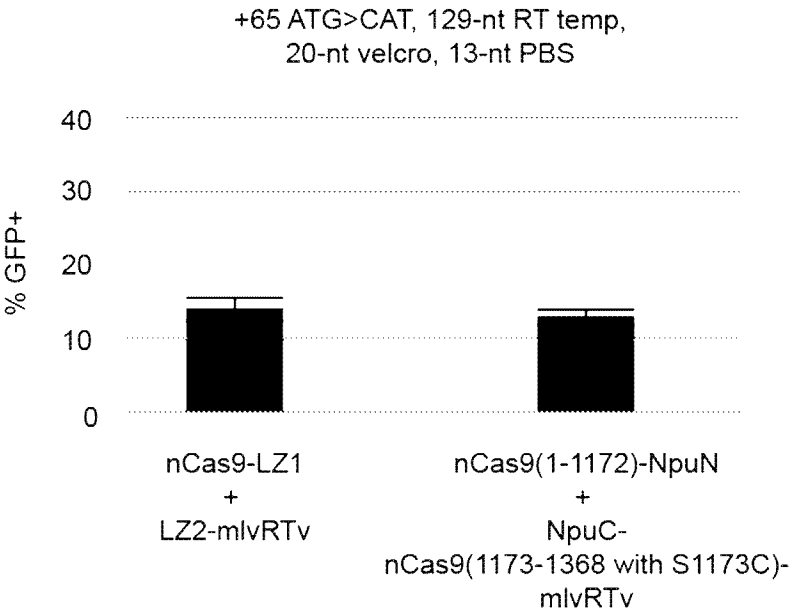


FIG. 17B

23/48

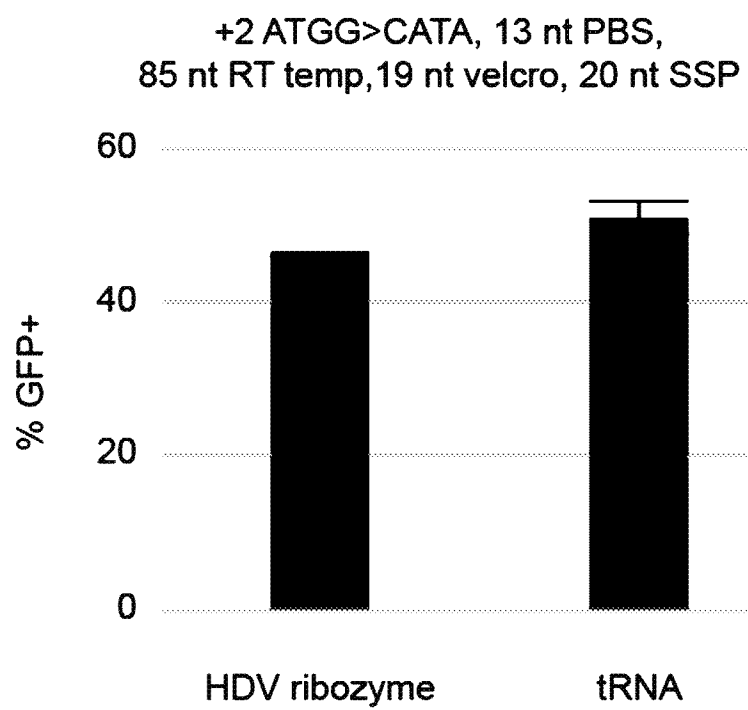


FIG. 18

FIG. 19A

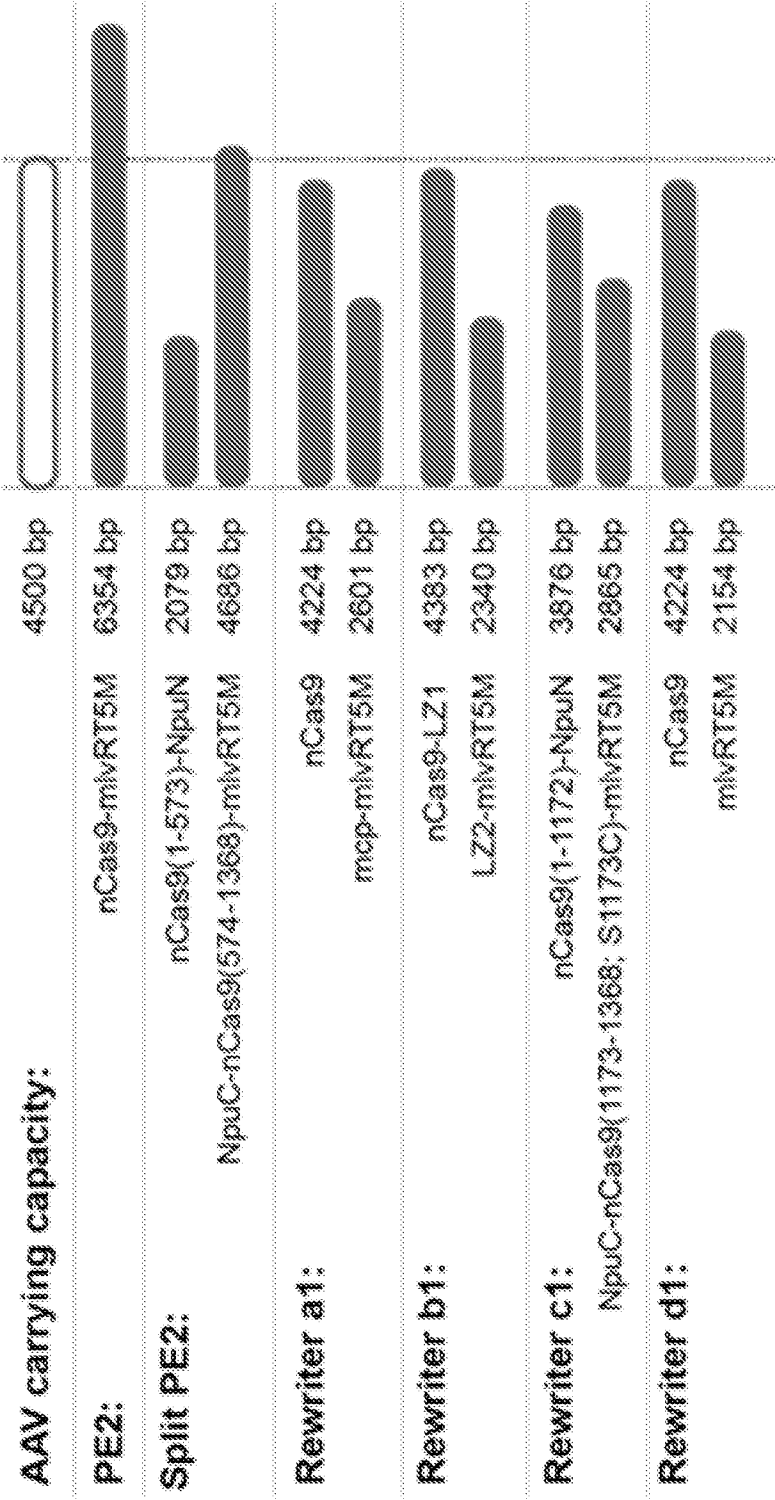


FIG. 19B

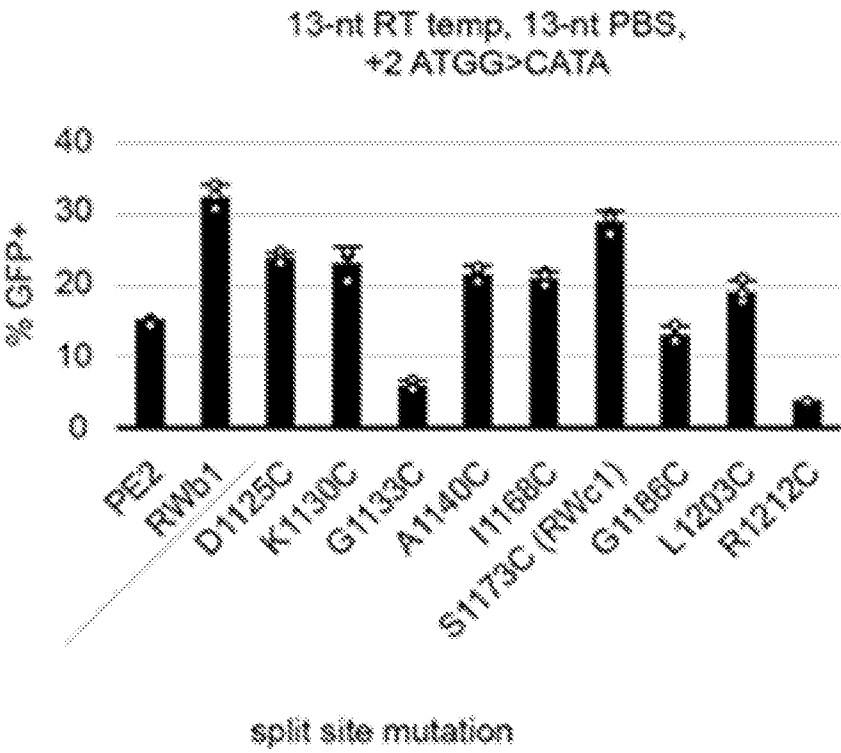


FIG. 19C

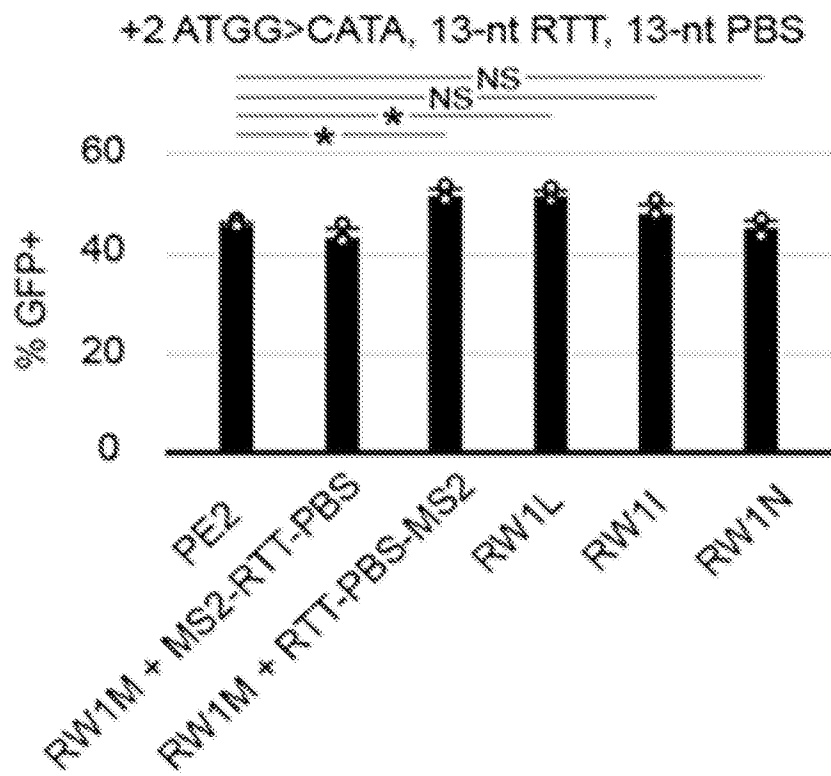


FIG. 19D

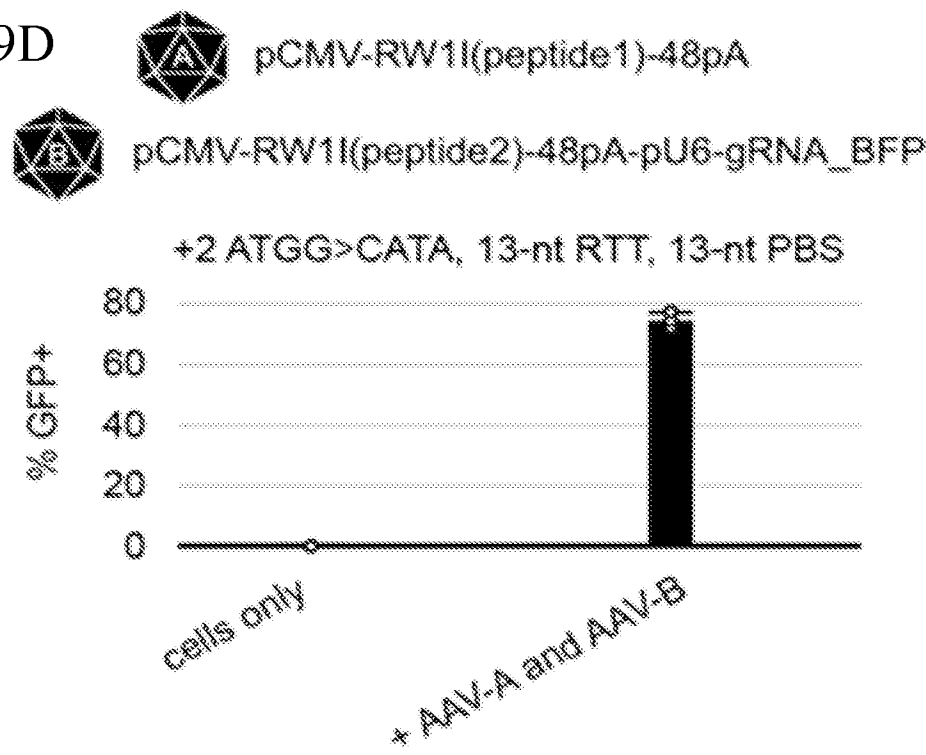


FIG. 20A



28/48

FIG. 20B

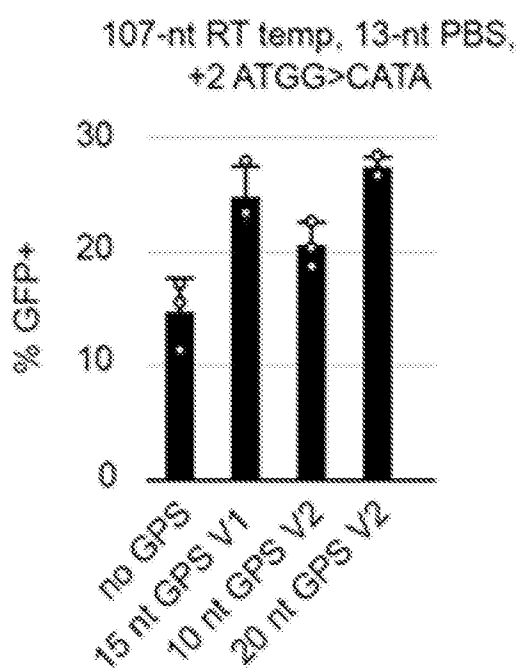
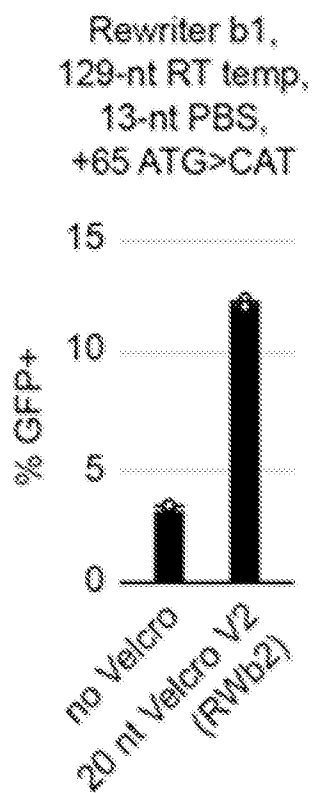


FIG. 20C



29/48

FIG. 21

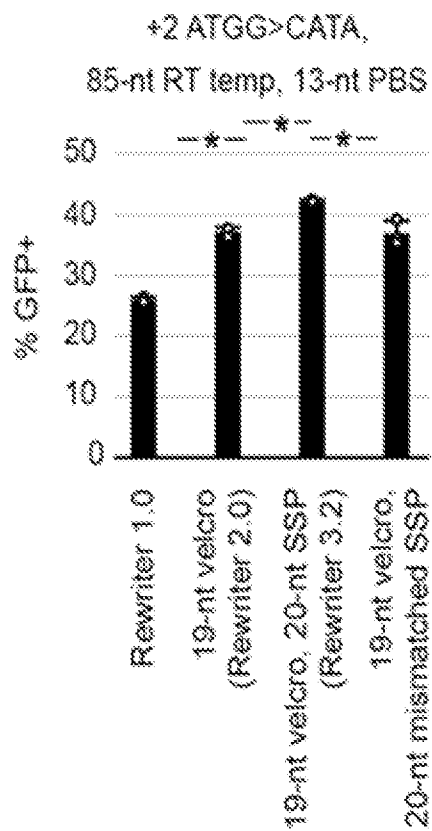


FIG. 22

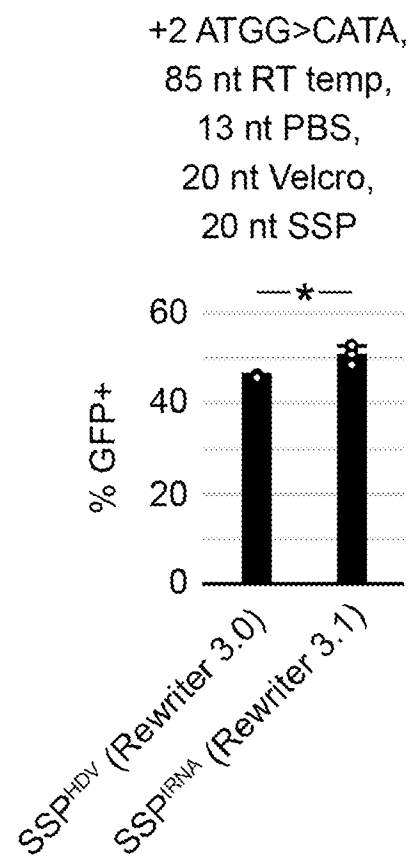


FIG. 23

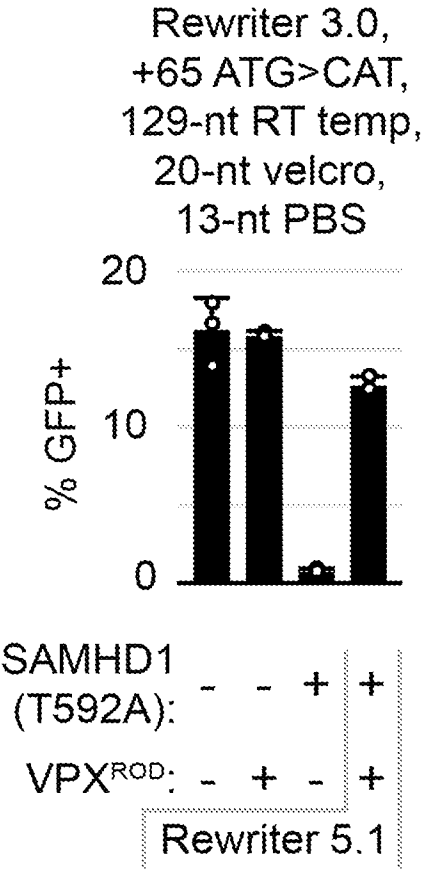


FIG. 24

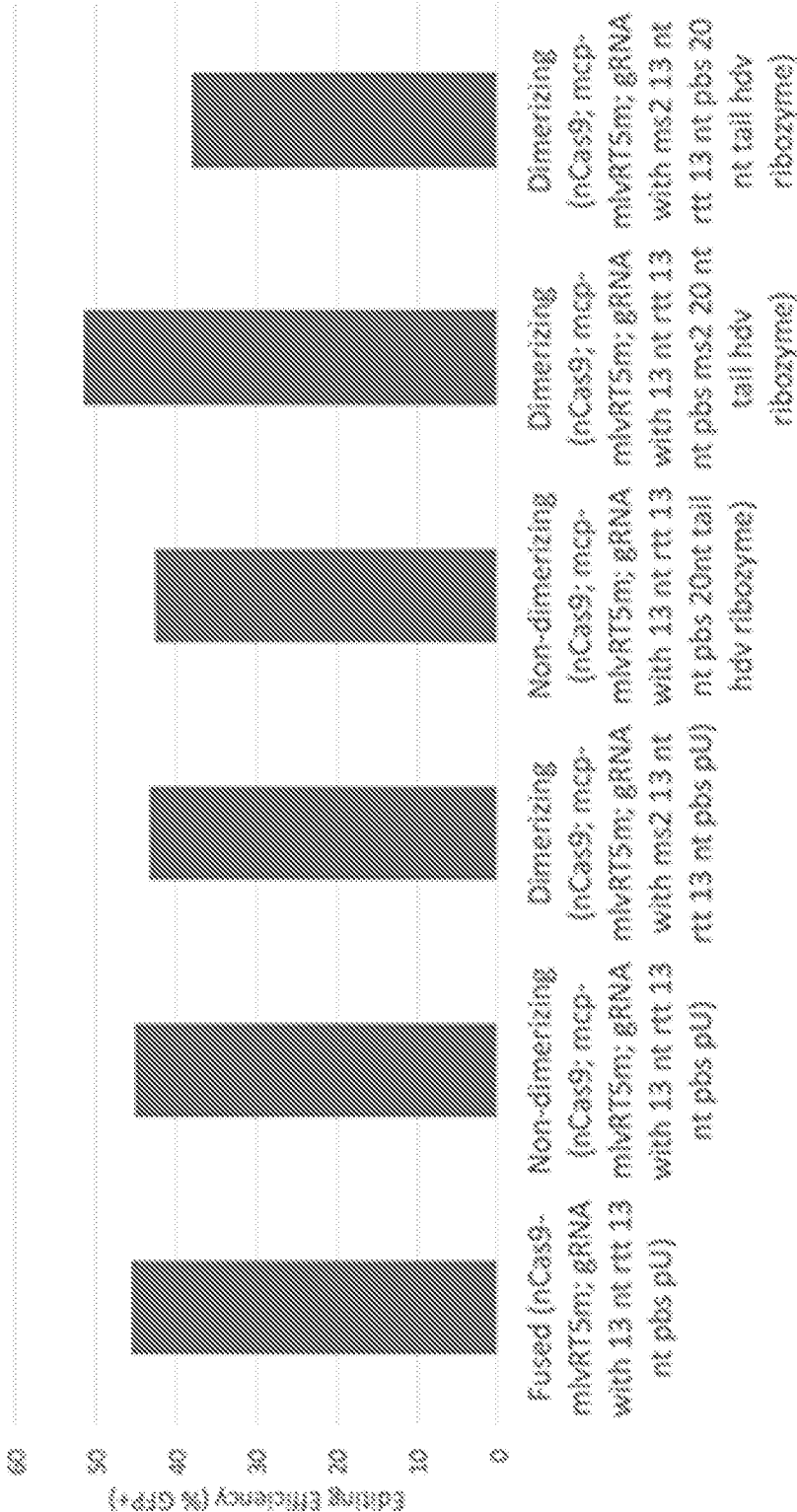


FIG. 25A

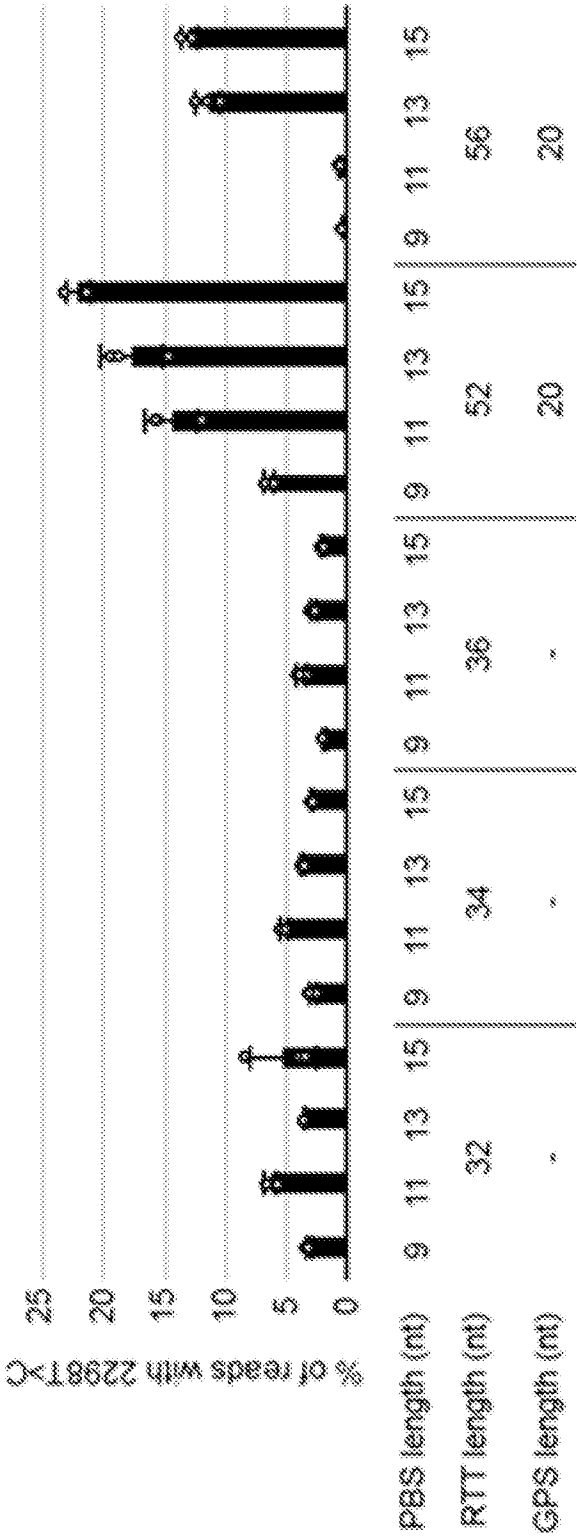


FIG. 25B

	<u>2298T>C</u>	<u>2316C>A</u>	<u>spacer</u>	
(SEQ ID NO: 120)	CTCTGGGCAGTGTGACTGC-AAAAAGAGGC-CAAGGACTTCAGTGTGACACCTGCAG			73.5% (wt)
(SEQ ID NO: 121)	CTCTGGGCAGTGTGACTGC-AAAAAGAGGC-CAAGGACTTCAGTGTGACACCTGCAG			18.8% (2298T>C, 2316C>A)
(SEQ ID NO: 122)	CTCTGGGCAGTGTGACTGC-AAAAAGAGGC-CAAGGACTTCAGTGTGACACCTGCAG			3.9% (2316C>T)
(SEQ ID NO: 123)	CTCTGGGCAGTGTGACTGC-AAAAAGAGGC-CAAGGACTTCAGTGTGACACCTGCAG			3.2% (2298T>C)
(SEQ ID NO: 124)	CTCTGGGCAGTGTGACTGC-AAAAAGAGGC-CAAGGACTTCAGTGTGACACCTGCAG			0.4% (2298T>C, 2315insA, 2316C>A)
(SEQ ID NO: 125)	CTCTGGGCAGTGTGACTGC-AAAAAGAGGC-CAAGGACTTCAGTGTGACACCTGCAG			0.2% (2298T>C, 2304insA, 2316C>A)

FIG. 25C

genome sequence with PAM bolded

(SEQ ID NO: 126) CTCTGGGCAGTGTGACTGC~~CA~~AAAAAGAG**CCCA**AGGACTTCAGTGTGACACCTGCAG
spacer

2316C>A is a silent mutation that was
encoded in the RTT to disrupt the PAM site
(SEQ ID NO: 127) CTCTGGGCAGTGTGACTGC~~CA~~AAAAAGAG**CCCA**AGGACTTCAGTGTGACACCTGCAG
2298T>C 2316C>A spacer

FIG. 25D

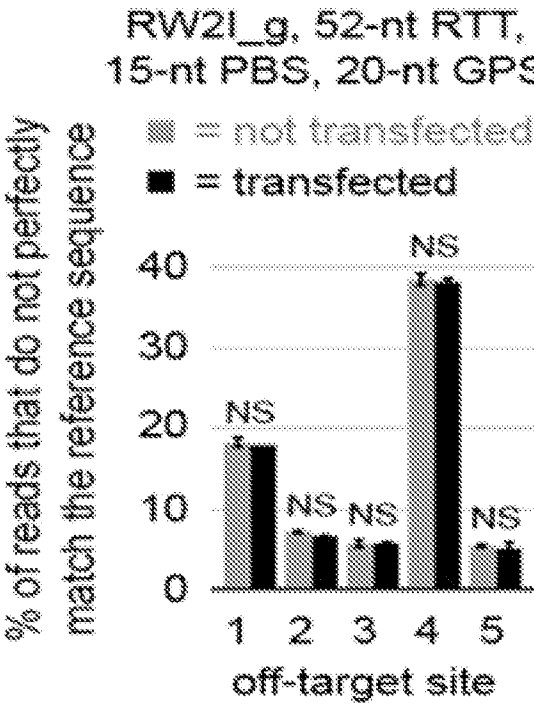
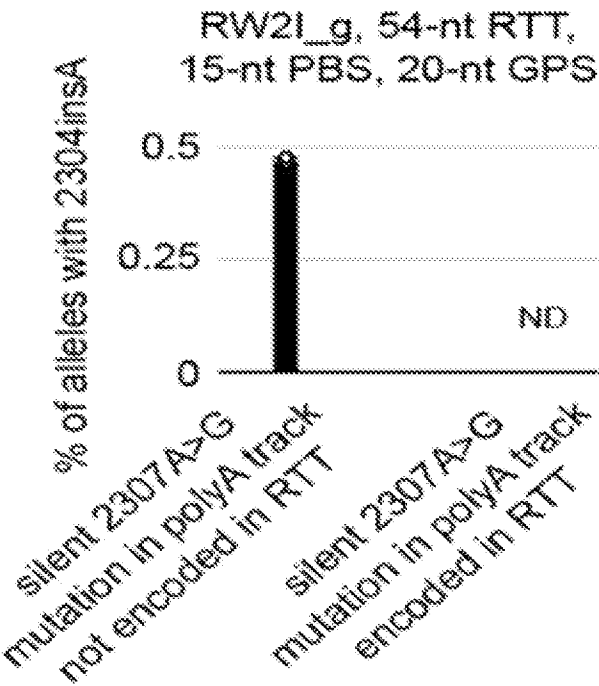


FIG. 25E



35/48

FIG. 25F

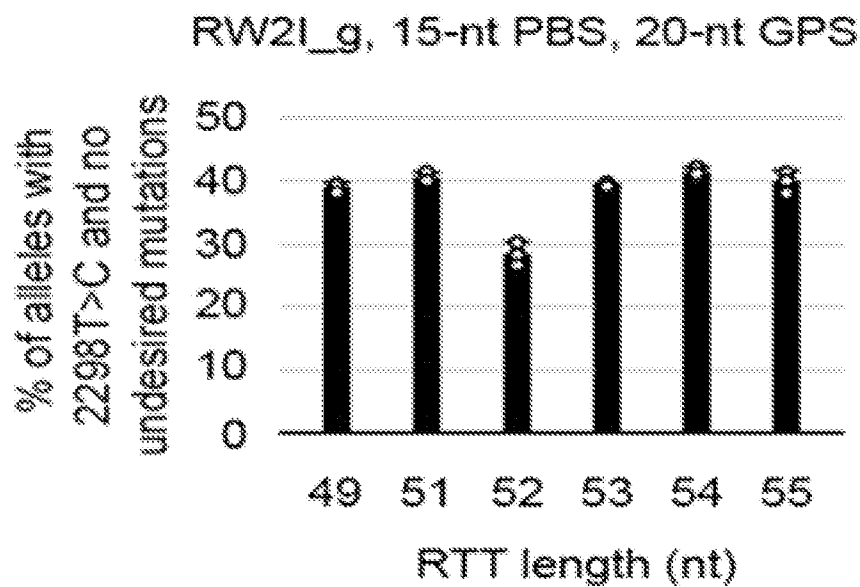


FIG. 25G

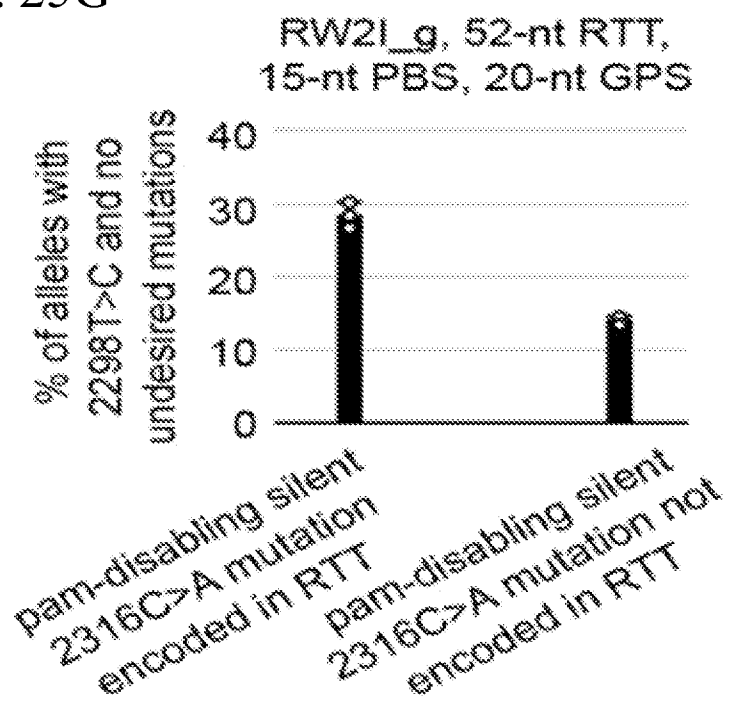
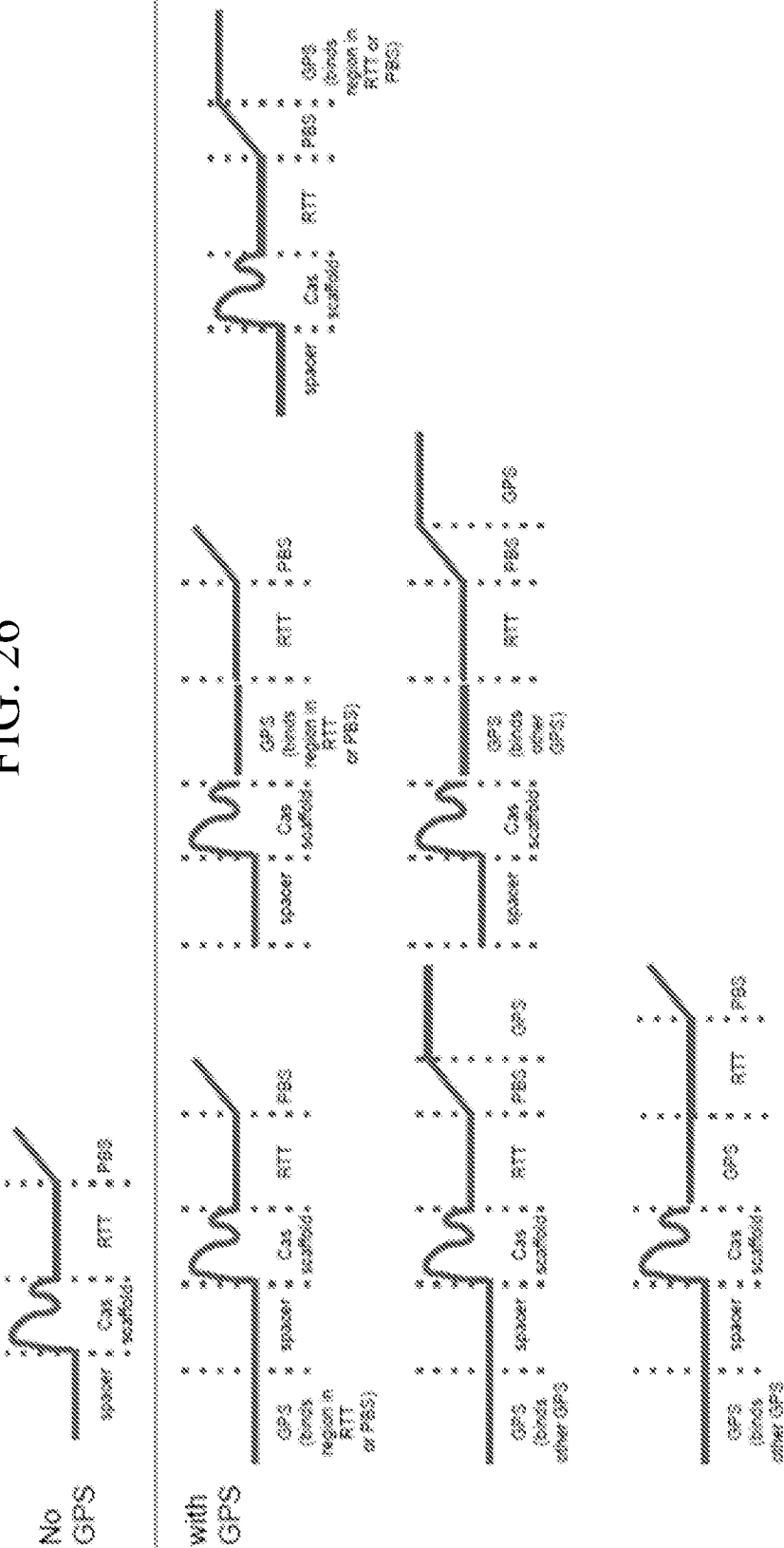


FIG. 26

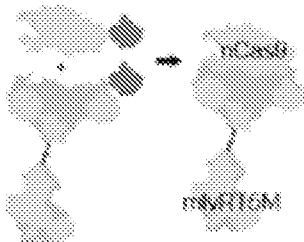


PE2: nCas9-mlvRT5M



FIG. 27

Split PE2: nCas9(1-573)-NpuN + NpuC-nCas9(574-1368)-mlvRT5M



RW1M: nCas9 + MBP-mlvRT5M



RW1L: nCas9-LZ1 + LZ2-mlvRT5M



RW1I: nCas9(1-1172)-NpuN + NpuC-nCas9(1173-1368; S1173C)-mlvRT5M



RW1N: nCas9 + mlvRT5M



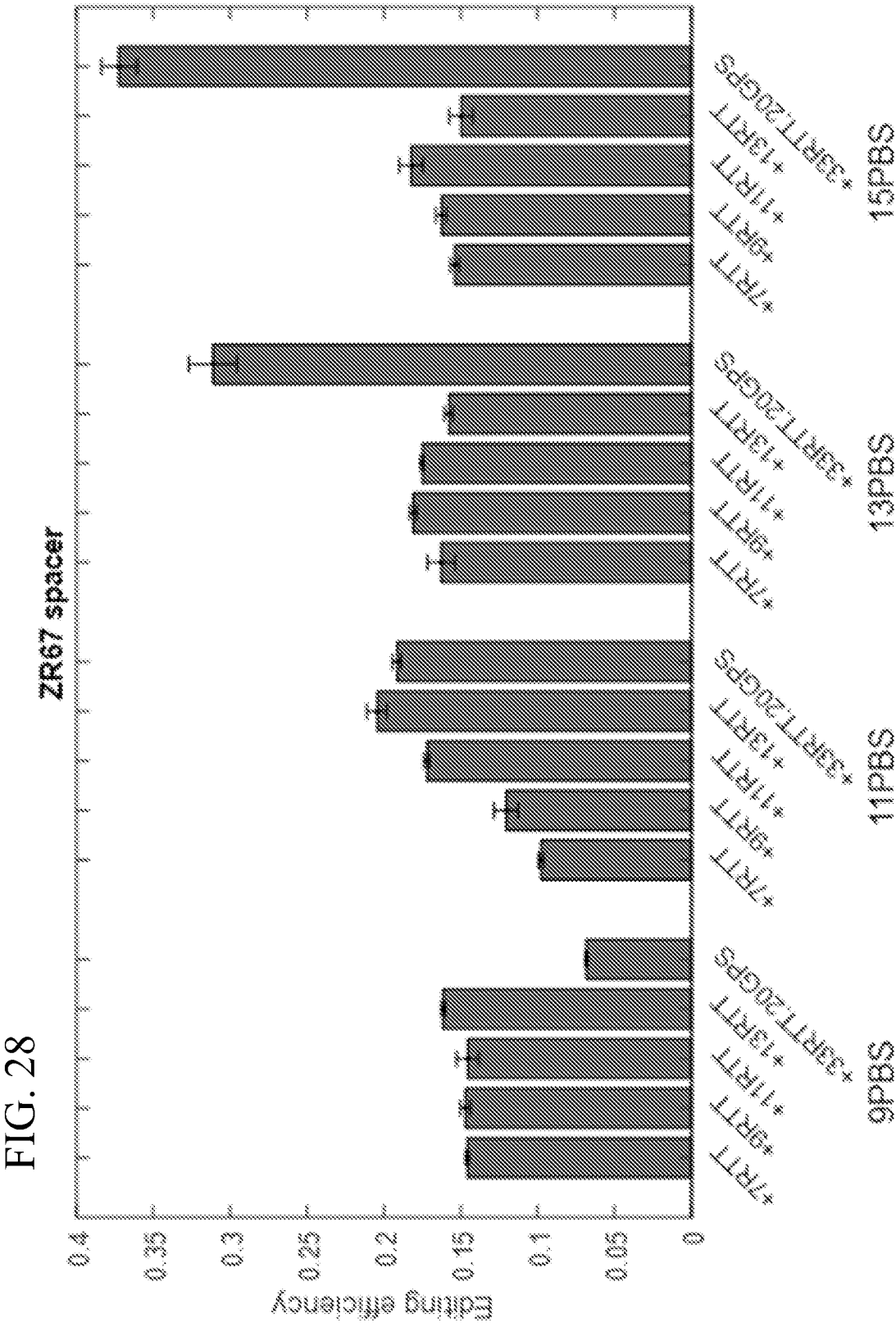


FIG. 29A

Ush2a editing with GGTGTCACACTGAAGTCCTT (SEQ ID NO: 128) spacer extended with 71ntRTT.15ntPBS.20ntGPS and AATTCTGCAATCCTCACTCT (SEQ ID NO: 129) spacer nicking the opposite strand

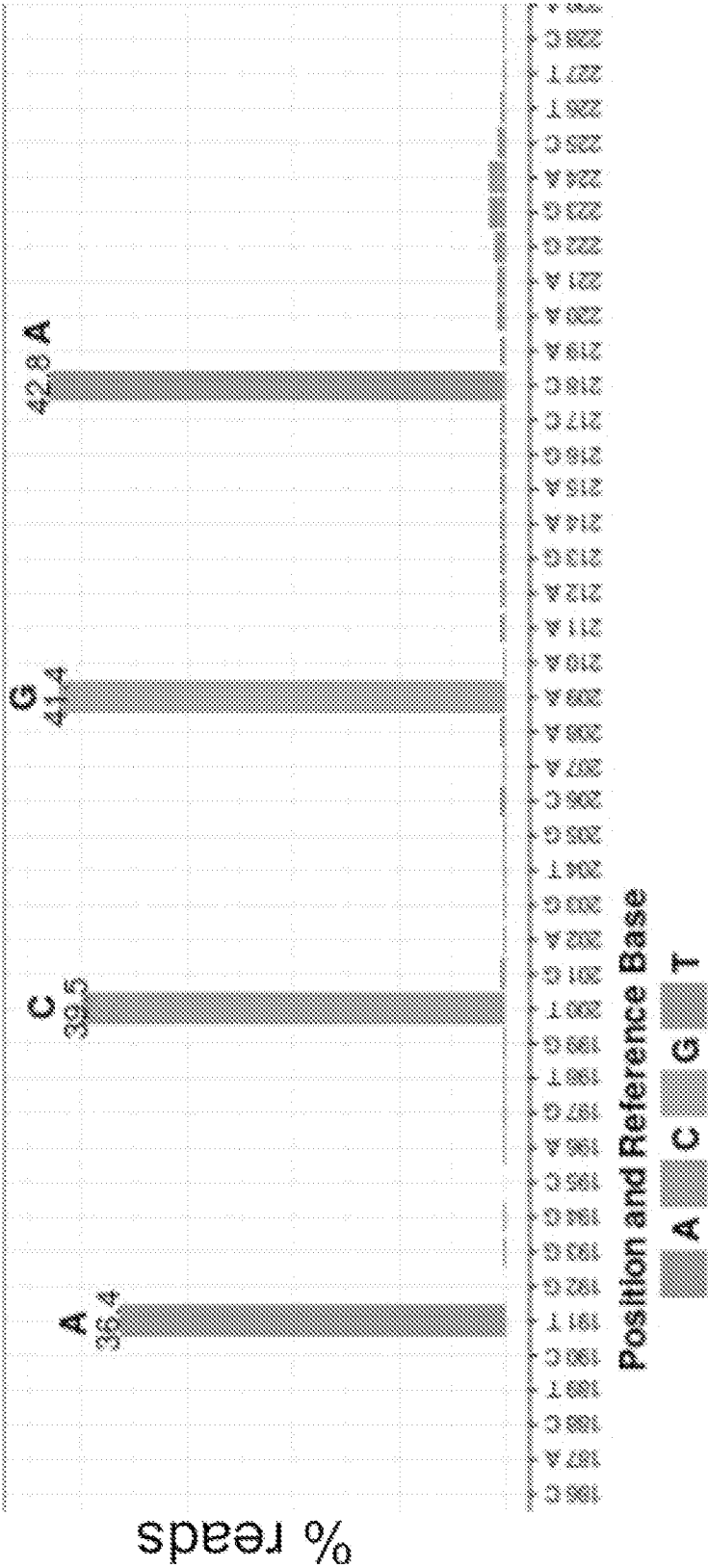


FIG. 29B

CFTR editing with GAAGGTGGAATCACACTGAG (SEQ ID NO: 130)
spacer extended with 48RTT.15ntPBS.20ntGPS

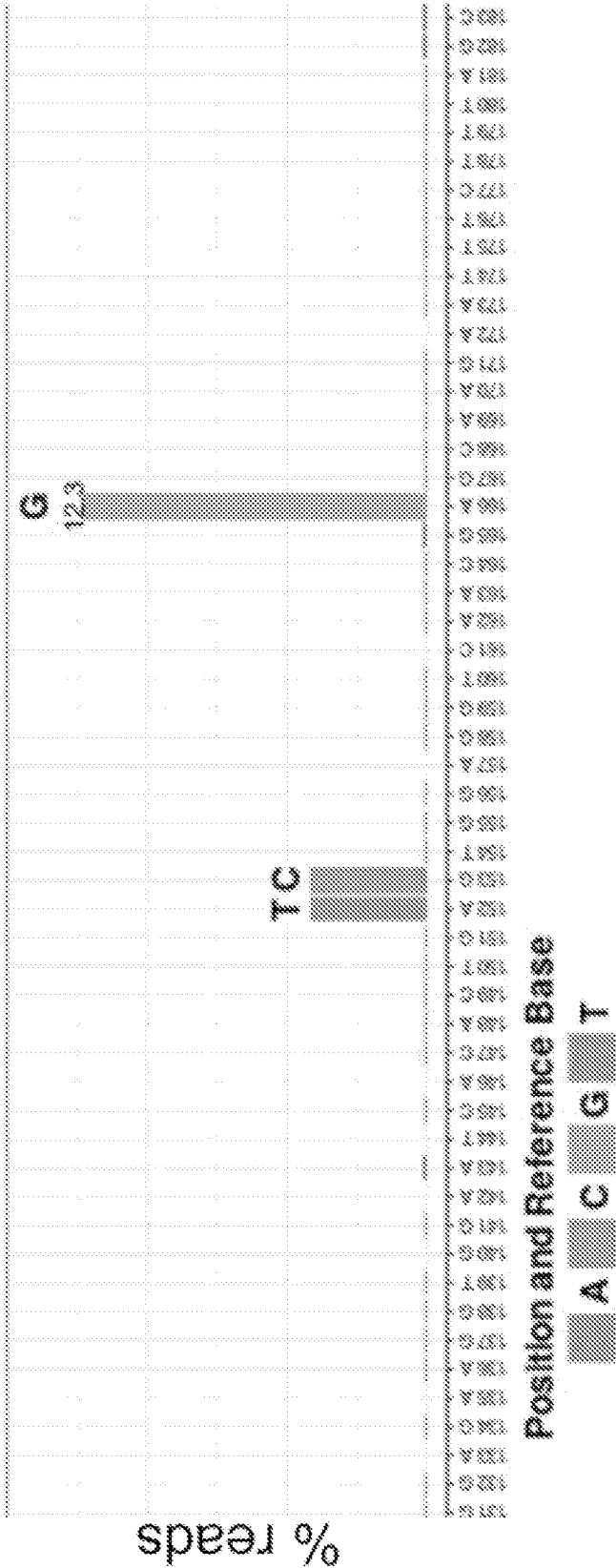


FIG. 29C

No transfection control for CFTR

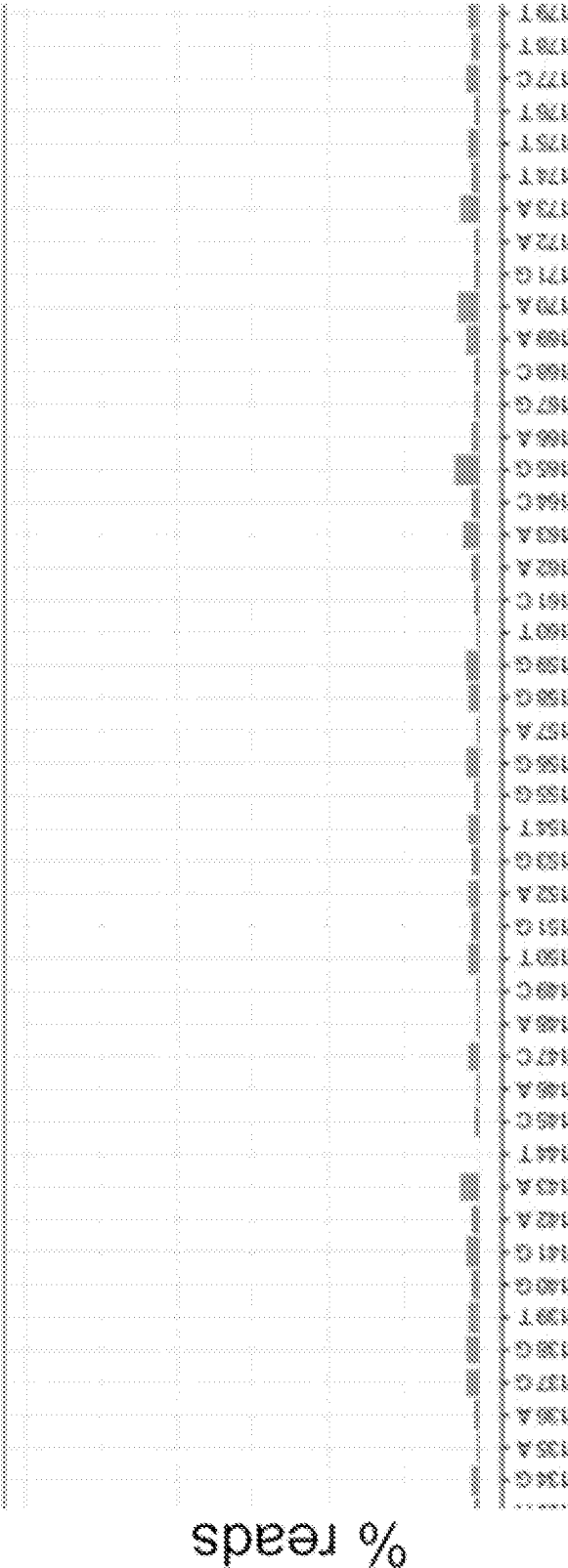


FIG. 29D

Abca4 editing with AAGAAAGTACCAAGGAAGTG (SEQ ID NO: 131) spacer extended with 55ntRTT.11ntPBS.20ntGPS

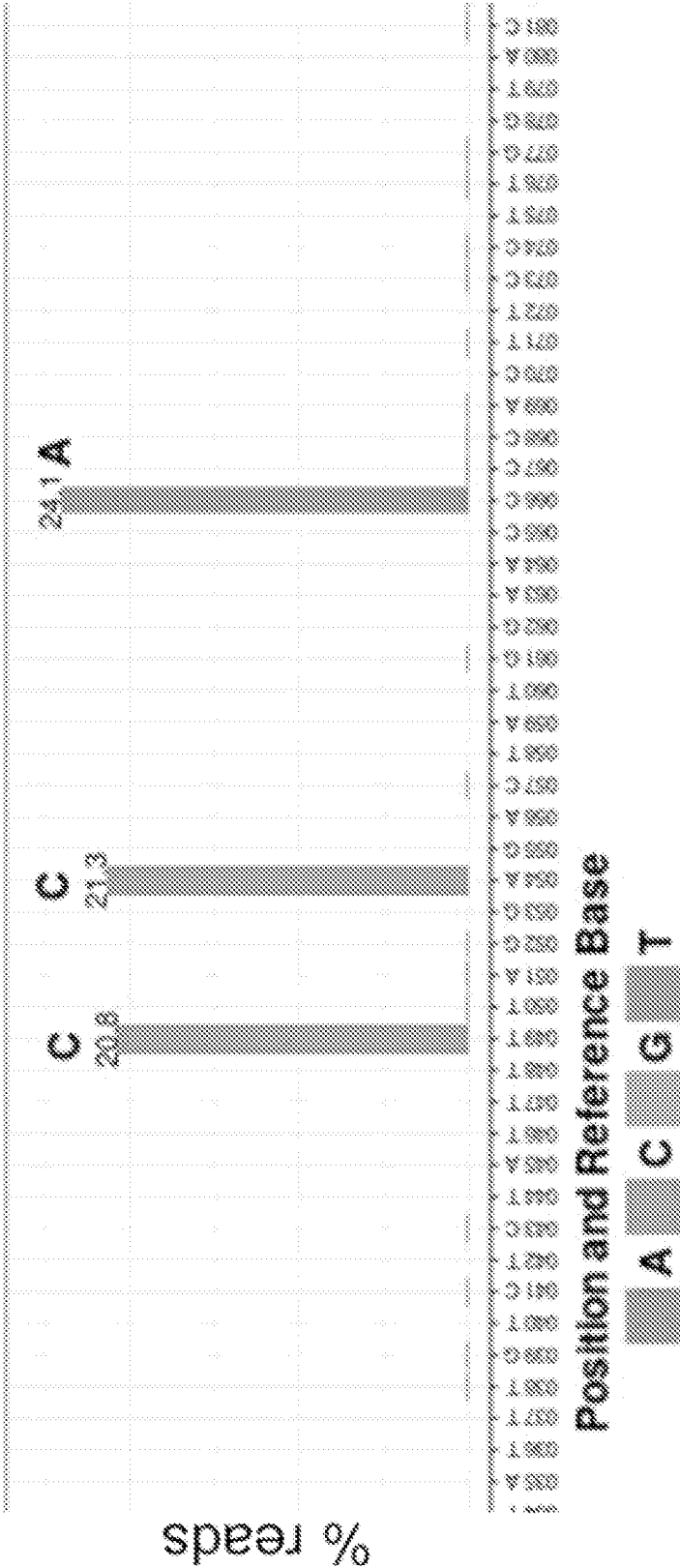


FIG. 29E

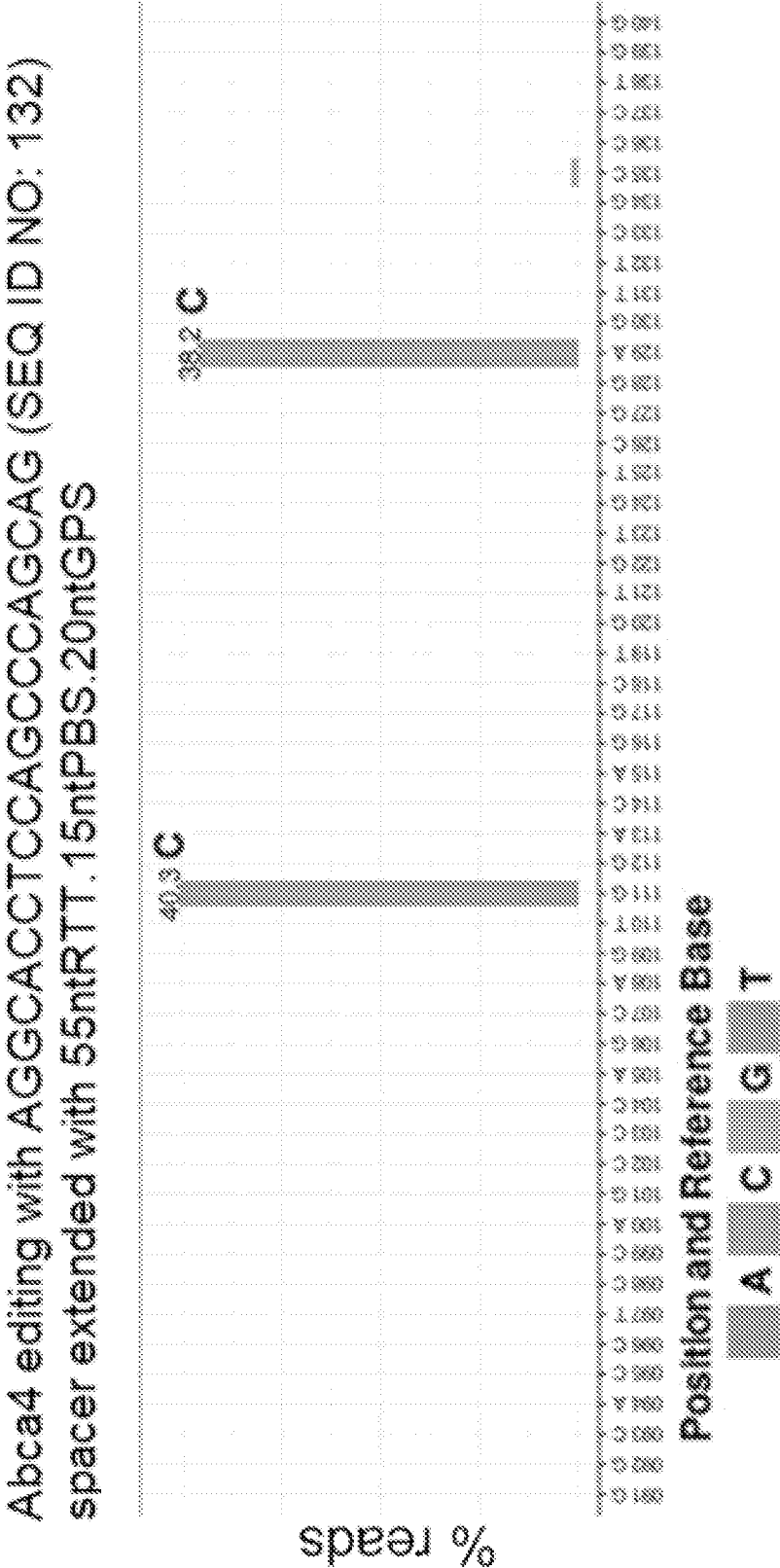
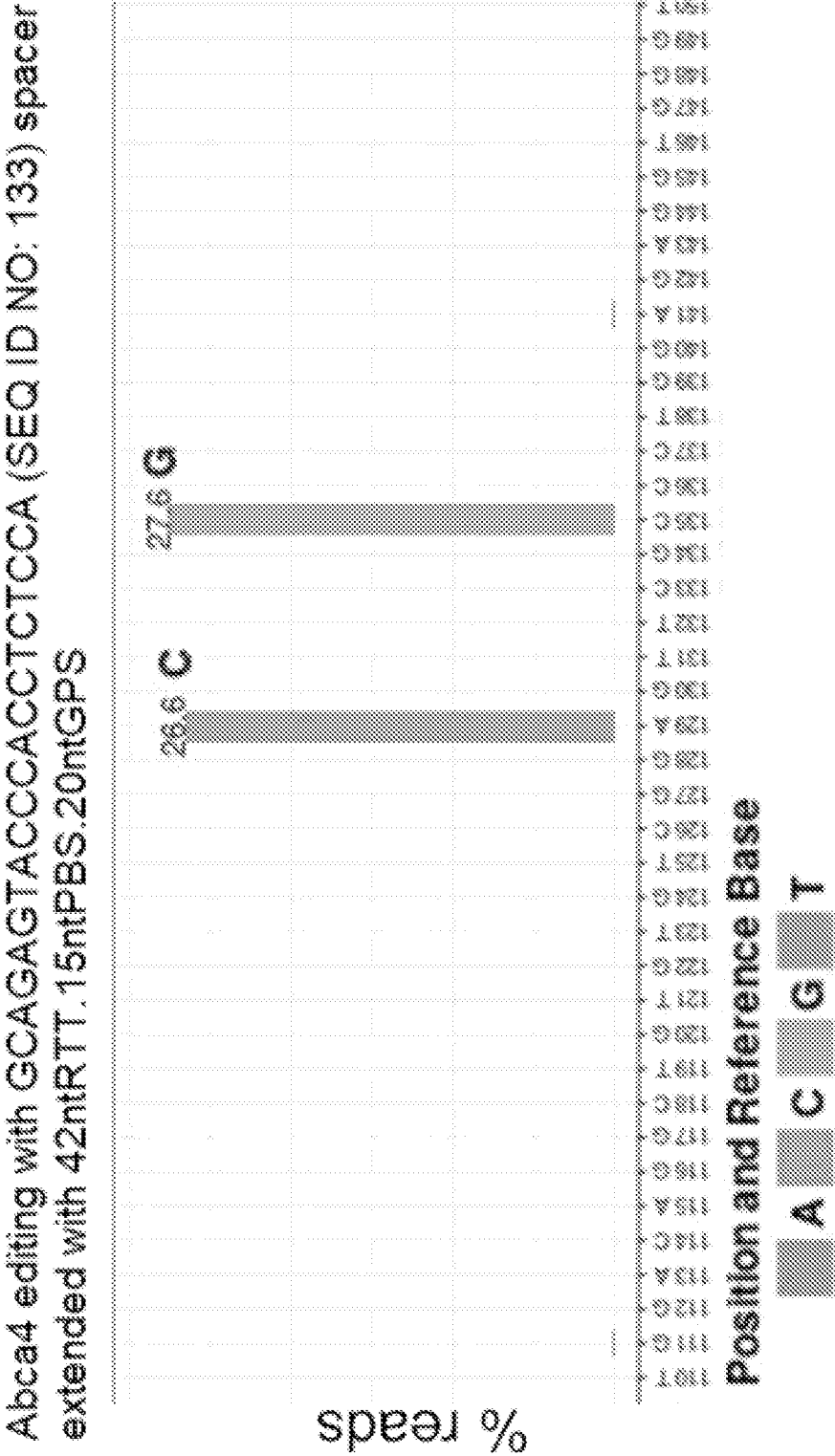


FIG. 29F



45/48

FIG. 30

reference

CTCTGGCAGTGTGAGTGCAAAAAGAGCCAAAGGACTTCAGTGTGACACCTGCAG
SEQ ID NO: 134

encoding the bolded 6 consecutive A's
in the RTT resulted in undesired edits

CTCTGGCAGTGCAGTGCAAAAAGAGC**AAAG** SEQ ID NO: 135

breaking up the track of 6 A's with a silent A to
mutation (bolded) eliminated undesired edits

CTCTGGCAGTGCAGTGCAAGAAAGAGC**AAAG** SEQ ID NO: 136

FIG. 31A

Cells treated with Rewriter HTT components including AGTCCTCAAGTCCTCCAG (SEQ ID NO: 137) spacer

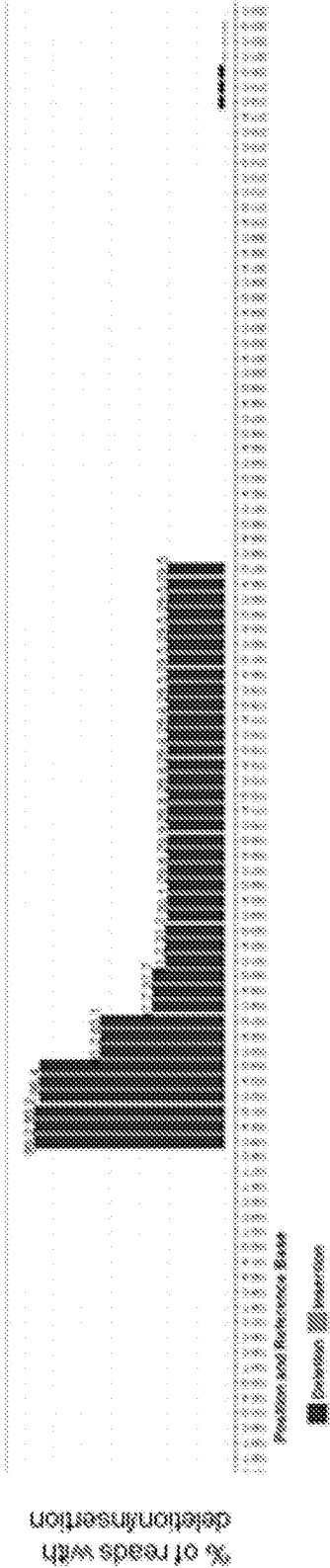


FIG. 31B Wildtype cells

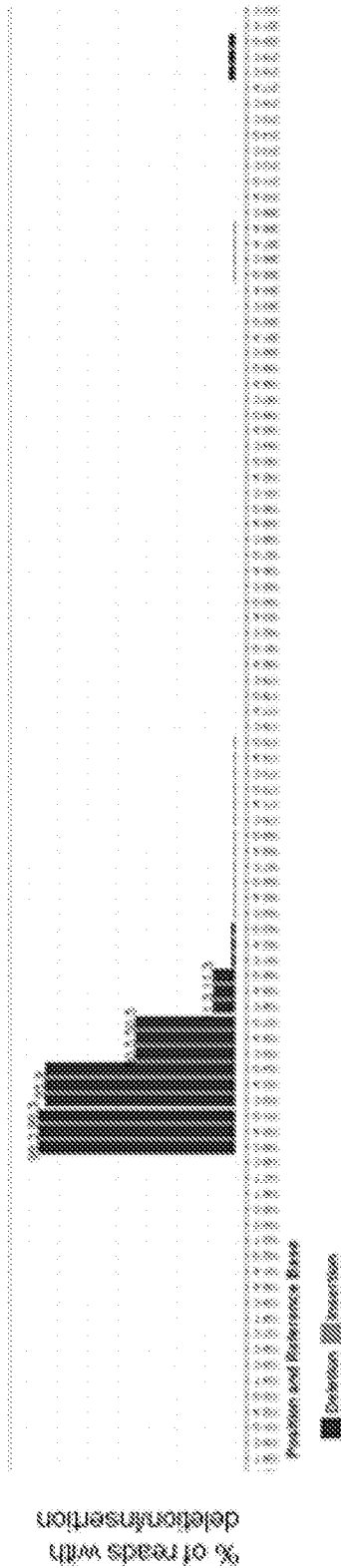
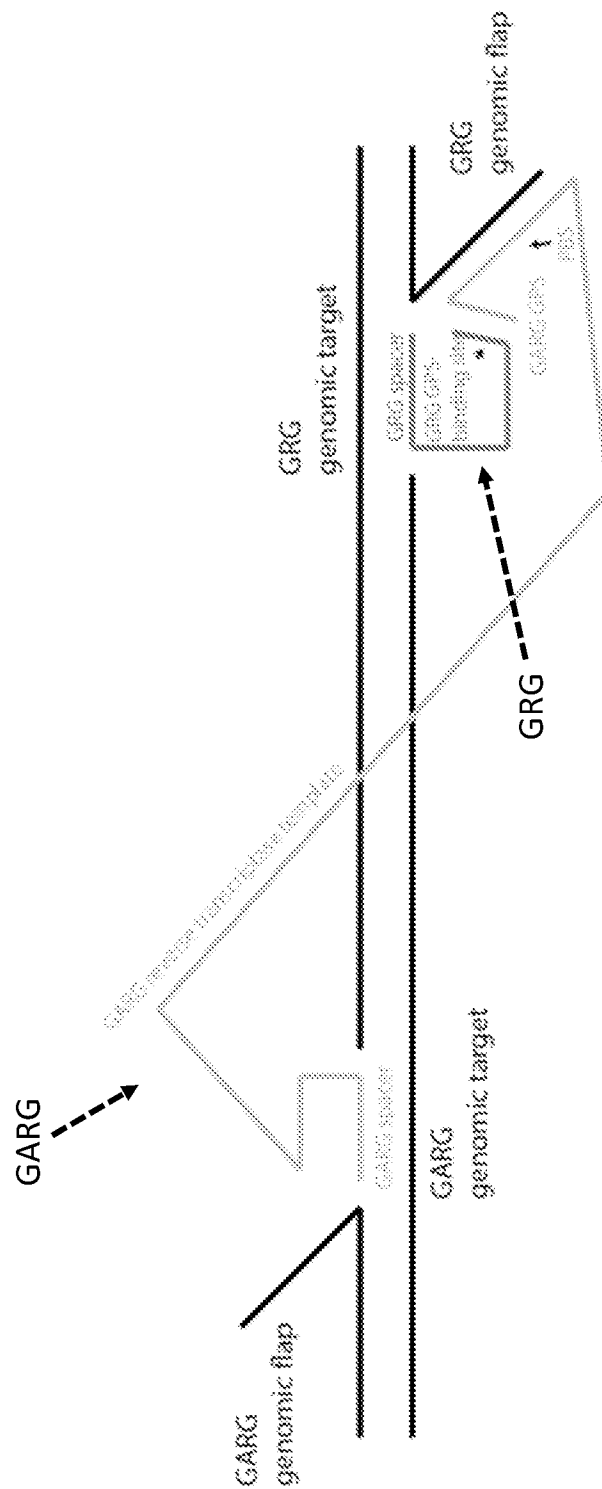
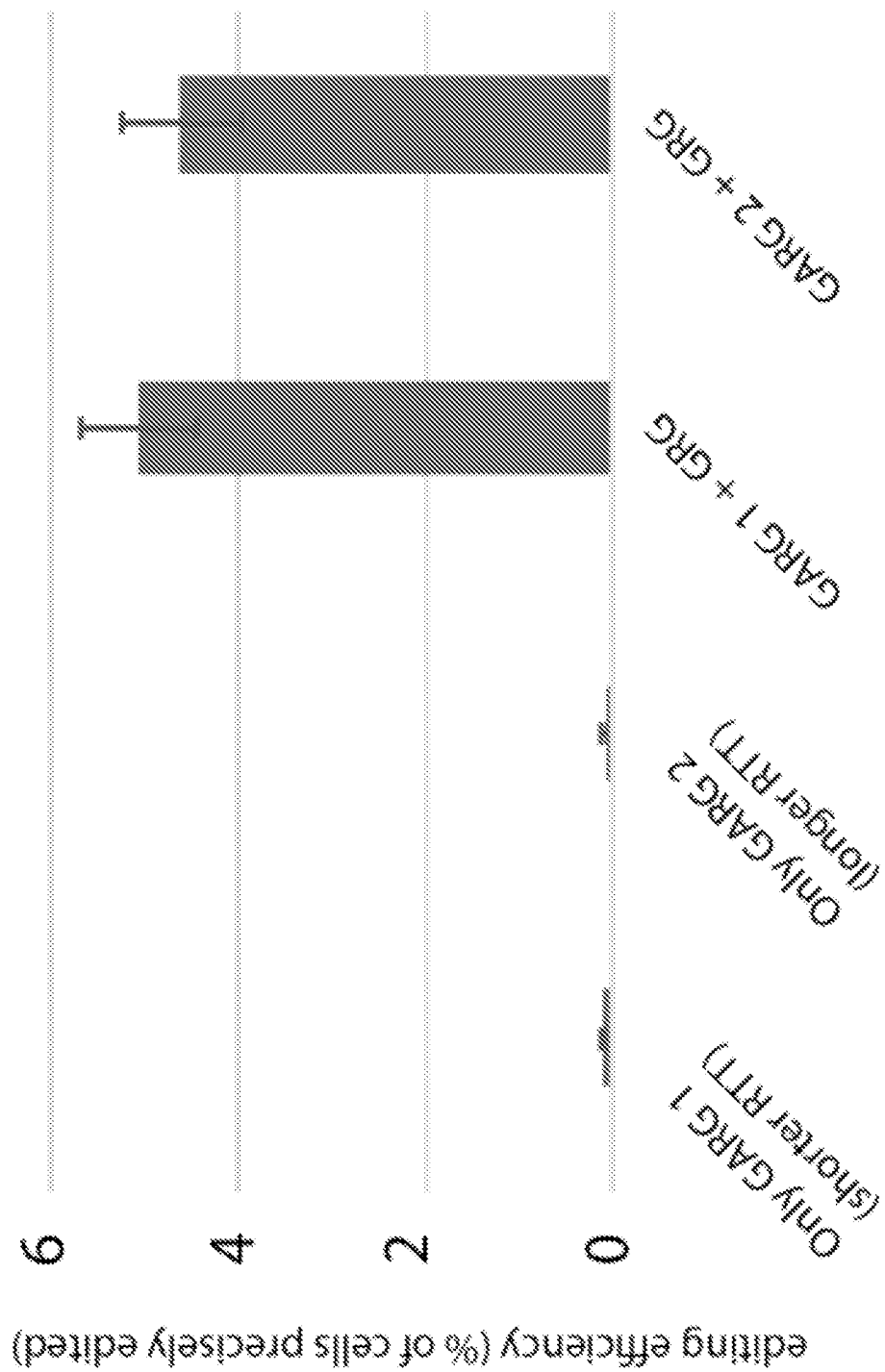


FIG. 32



48/48

FIG. 33



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/023041

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

SEQ ID NOs: 3-22, 40-80, and 95 were searched.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/023041

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 25-27, 29-32, 38-47, 51-77, 79, 89-99, 107-113, 115-123
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/023041

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 9/12; C12N 9/22 (2021.01)

CPC - C12N 9/1276; C12N 9/22; C12N 2310/20; C12Y 207/07049 (2021.05)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2019/0350978 A1 (JUNO THERAPEUTICS, INC.) 21 November 2019 (21.11.2019) entire document	1-4, 14-20, 78, 80-82 5-13, 83, 100-106, 124, 125
X -- Y	WO 2020/047124 A1 (FLAGSHIP PIONEERING, INC.) 05 March 2020 (05.03.2020) entire document	21, 24, 28, 33 22, 23, 34, 35
X	WO 2020/012335 A1 (ALIA THERAPEUTICS S.R.L.) 16 January 2020 (16.01.2020) entire document	36
X	WO 2018/049168 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 15 March 2018 (15.03.2018) entire document	37, 114
X	US 2018/0237770 A1 (CARIBOU BIOSCIENCES, INC.) 23 August 2018 (23.08.2018) entire document	48-50, 127-131
X	WO 2020/051561 A1 (BEAM THERAPEUTICS INC.) 12 March 2020 (12.03.2020) entire document	84-88
Y	ZHANG et al. "Conserved Herpesvirus Protein Kinases Target SAMHD1 to Facilitate Virus Replication," Cell Reports, 09 July 2019 (09.07.2019), Vol. 28, Iss. 2, Pgs. 449-459.e5. entire document	5



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 June 2021

Date of mailing of the international search report

JUL 13 2021

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Harry Kim

Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/023041

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2018/0313843 A1 (JOHANN WOLFGANG GOETHE-UNIVERSITÄT FRANKFURT AM MAIN) 01 November 2018 (01.11.2018) entire document	6, 7, 83
Y	US 2020/0038424 A1 (THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK et al) 06 February 2020 (06.02.2020) entire document	8-10
Y	US 2002/0192788 A1 (IHLENFELDT et al) 19 December 2002 (19.12.2002) entire document	11-13
Y	US 2018/0230494 A1 (THE GENERAL HOSPITAL CORPORATION) 16 August 2018 (16.08.2018) entire document	22, 23
Y	WO 2019/070843 A1 (BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM) 11 April 2019 (11.04.2019) entire document	34, 35, 126
Y	US 2018/0298414 A1 (THERMO FISHER SCIENTIFIC BALTICS UAB) 18 October 2018 (18.10.2018) entire document	100-106, 124, 125
Y	PALIKSA et al. "Decreased Km to dNTPs is an essential M-MuLV reverse transcriptase adoption required to perform efficient cDNA synthesis in One-Step RT-PCR assay," Protein Engineering Design & Selection, 15 March 2018 (15.03.2018), Vol. 31, Iss. 3, Pgs. 79-89. entire document	105
Y	US 2014/0010839 A1 (BENKIRANE et al) 09 January 2014 (09.01.2014) entire document	126
P, X	WO 2020/191153 A2 (THE BROAD INSTITUTE, INC. et al) 24 September 2020 (24.09.2020) entire document	1-24, 28, 33-37, 48-50, 78, 80-88, 100-106, 114, 124-131
A	ANZALONE et al. "Search-and-replace genome editing without double-strand breaks or donor DNA," Nature, 21 October 2019 (21.10.2019), Vol. 576, No. 7785, Pgs. 149-157. entire document	1-24, 28, 33-37, 48-50, 78, 80-88, 100-106, 114, 124-131
A	ST GELAIS et al. "SAMHD1 restricts HIV-1 infection in dendritic cells (DCs) by dNTP depletion, but its expression in DCs and primary CD4+ T-lymphocytes cannot be upregulated by interferons," Retrovirology, 11 December 2012 (11.12.2012), Vol. 9:105, Pgs. 1-15. entire document	1-24, 28, 33-37, 48-50, 78, 80-88, 100-106, 114, 124-131
A	US 2019/0062409 A1 (UNIVERSITY OF MASSACHUSETTS) 28 February 2019 (28.02.2019) entire document	1-24, 28, 33-37, 48-50, 78, 80-88, 100-106, 114, 124-131