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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 November 2019 (14.11.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/213776 A1

(51) International Patent Classification:

C12N 15/11 (2006.01)	*C12N 15/55* (2006.01)
A61K 38/47 (2006.01)	*C12N 15/63* (2006.01)
A61K 48/00 (2006.01)	*C12N 15/864* (2006.01)
A61P 1/16 (2006.01)	*C12N 15/90* (2006.01)
A61P 3/00 (2006.01)	*C12N 9/22* (2006.01)
C12N 15/09 (2006.01)	*C07H 21/02* (2006.01)

(21) International Application Number:
PCT/CA2019/050629

(22) International Filing Date:
10 May 2019 (10.05.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/670,135 11 May 2018 (11.05.2018) US

(71) Applicant: **UNIVERSITÉ LAVAL** [CA/CA]; Pavillon des sciences de l'éducation, 2320 rue des Bibliothèques, local 1434, Québec, Québec G1V 0A6 (CA).

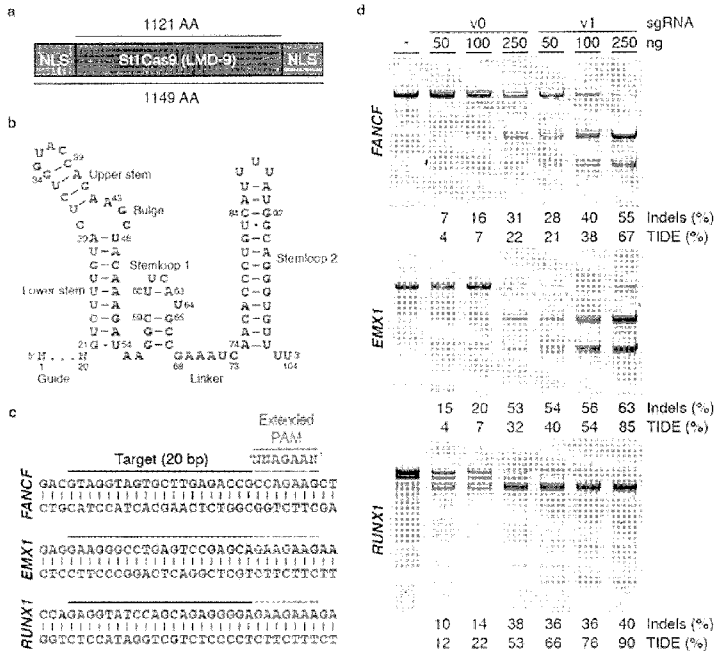
(72) Inventor: **DOYON, Yannick**; 1282 Roland-Desmeules, Québec, Québec G1X 4Y3 (CA).

(74) Agent: **LAVERY, DE BILLY, L.L.P.**; 1, Place Ville Marie, Bureau 4000, Montréal, Québec H3B 4M4 (CA).

(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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

(54) Title: CRISPR/CAS9 SYSTEM AND USES THEREOF



FIGs. 1a-1d

(57) Abstract: Methods and products are described herein for the modification of nucleic acids using a CRISPR/Cas9 system. Also described herein are uses of such methods and products for the modification of a target nucleic acid in a cell, *in vitro* or in vivo. Such methods and products may also be used for prevention or treatment of a condition associated with a target polynucleotide.

[Continued on next page]

WO 2019/213776 A1

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))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

CRISPR/CAS9 SYSTEM AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of U.S. Provisional Application serial no. 62/670,135 filed on May 11, 2018, which is incorporated herein by reference in its entirety.

SEQUENCE LISTING

This application contains a Sequence Listing in computer readable form entitled "G11229_398_SeqList.txt", created on May 9, 2019 and having a size of about 343 KB. The computer readable form is incorporated herein by reference in its entirety.

TECHNICAL FIELD

The present disclosure generally relates to the modification of nucleic acids, and more particularly to the to the modification of nucleic acids using a CRISPR/Cas9 system.

BACKGROUND ART

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins are components of prokaryotic adaptive immune systems that have been harnessed for robust genome editing[1]. Type II-based tools rely on a large multidomain endonuclease, Cas9, guided to its DNA target by an engineered single-guide RNA (sgRNA) chimera[2,3,4]. The Cas9-sgRNA binary complex finds its target through recognition of a short sequence called the protospacer adjacent motif (PAM) and subsequent base pairing of the guide RNA with the DNA to generate a specific double-strand break (DSB)[1,5]. While *Streptococcus pyogenes* (SpCas9) remains the most widely used Cas9 variant for genome engineering, other RNA-guided nucleases have also been identified[4,6]. However, certain bacterial CRISPR/Cas enzymes were found to be inactive in human cells despite being accurately reprogrammed for DNA binding and cleavage *in vitro*[7-10]. An even greater challenge has been implementation *in vivo*, examples including the use of the type II-A Cas9 from *Staphylococcus aureus* (SaCas9) for *in vivo* editing using recombinant Adeno-Associated Virus (rAAV) vectors[7,11,12], as well as Cas9s from *Campylobacter jejuni* and *Neisseria meningitidis*[13-15].

In vivo genome editing offers the possibility to generate phenotypes in animal models in order to better recapitulate the interactions between cell types and organs. In addition, it can be contemplated as a novel class of human therapeutics that enables precise molecular correction of genetic defects underlying diseases. As such, it has for example been shown that rAAV- and zinc-finger nuclease (ZFN)-mediated liver targeting can correct disease phenotypes in neonatal and adult mouse models, a process currently under clinical investigation[16-19].

There is therefore a need for further development of robust and wide-ranging CRISPR-based technologies, for example for *in vivo* editing.

The present description refers to a number of documents, the content of which is herein incorporated by reference in their entirety.

SUMMARY OF THE DISCLOSURE

The present disclosure generally relates to the modification of nucleic acids, and more particularly to the to the modification of nucleic acids using a CRISPR/Cas9 system. Methods and products are described herein for the modification of nucleic acids using a CRISPR/Cas9 system. Also described herein are uses of such methods and products for the modification of a target nucleic acid in a cell, *in vitro* or *in vivo*. Such methods and products may also be used for prevention or treatment of a condition associated with a target polynucleotide.

In various aspects and embodiments, the present disclosure provides the following items 1 to 136:

1. A sgRNA for modification of a target polynucleotide in a cell, comprising:

(a) a guide segment comprising a guide sequence corresponding to a region of the target polynucleotide;

(b) a first hairpin-forming segment located 3' to the guide sequence, the first hairpin hairpin-forming segment being capable of forming a hairpin comprising a stem portion and a loop portion, wherein the stem portion does not comprise a sequence corresponding to an RNA polymerase III termination signal.

2. The sgRNA of item 1, wherein the stem portion does not comprise more than 4 consecutive uracil nucleotides.

3. The sgRNA of item 1, wherein the stem portion does not comprise more than 3 consecutive uracil nucleotides.

4. The sgRNA of item 1, wherein the stem portion comprises a first stem portion and a second stem portion, wherein the first stem portion does not comprise a sequence corresponding to an RNA polymerase III termination signal.

5. The sgRNA of item 4, wherein the first stem portion does not comprise more than 4 consecutive uracil nucleotides.

6. The sgRNA of item 5, wherein the first stem portion does not comprise more than 3 consecutive uracil nucleotides.

7. The sgRNA of any one of items 4 to 6, wherein the first stem portion and second stem portion are separated by a first bulge portion.

8. The sgRNA of any one of items 1 to 8, wherein the loop portion comprises or consists of a sequence of 3 to 6 nucleotides.

9. The sgRNA of item 8, wherein the loop portion comprises or consists of a sequence of 3 to 5 nucleotides.

10. The sgRNA of item 9, wherein the loop portion comprises or consists of a sequence of 4 nucleotides.

11. The sgRNA of item 10, wherein the loop portion comprises or consists of the nucleotide sequence $N^1N^2N^3N^4$, wherein N^1, N^2, and N^3 are each independently A, C, G or U, and N^4 is C or G.

12. The sgRNA of item 10 or 11, wherein the loop portion comprises or consists of the nucleotide sequence $N^1N^2N^3N^4$, wherein N^1, N^3, and N^4 are each independently A, C, G or U, and N^2 is U, G or A.

13. The sgRNA of item 11 or 12, wherein N^1 is G.

14. The sgRNA of any one of items 11 to 13, wherein N^2 is U.

15. The sgRNA of any one of items 11 to 14, wherein N^3 is A.

16. The sgRNA of any one of items 11 to 15, wherein N^4 is C.

17. The sgRNA of any one of items 10 to 16, wherein the loop portion comprises or consists of the nucleotide sequence GUAC.

18. The sgRNA of any one of items 4 to 17, wherein the second stem portion comprises or consists of a hybrid of 4 nucleotide pairs.

19. The sgRNA of item 18, wherein the fourth pair of the hybrid of the second stem portion, distal to the first stem portion, is a G-C pair.

20. The sgRNA of item 18 or 19, wherein the hybrid of the second stem portion comprises or consists of the sequence 5'-UCUG-3' hybridized to the sequence 5'-CAGA-3'.

21. The sgRNA of any one of items 4 to 20, wherein the first stem portion comprises or consists of a hybrid of at least 5 nucleotide pairs.

22. The sgRNA of any one of items 4 to 21, wherein the first stem portion comprises or consists of a hybrid of not more than 12 nucleotide pairs.

23. The sgRNA of item 21 or 22, wherein the first stem portion comprises or consists of a hybrid of 6 to 10 nucleotide pairs.

24. The sgRNA of item 23, wherein the first stem portion comprises or consists of a hybrid of 7 to 9 nucleotide pairs.

25. The sgRNA of item 24, wherein the first stem portion comprises or consists of a hybrid of 8 nucleotide pairs.

26. The sgRNA of any one of items 4 to 25, wherein the first stem portion does not comprise a mismatch.

27. The sgRNA of any one of items 4 to 26, wherein the hybrid of the first stem portion comprises or consists of the sequence 5'-UCUUUGUA-3' hybridized to the sequence 5'-UACAAAGA-3'.

28. The sgRNA of any one of items 4 to 24, wherein the first stem portion comprises a single mismatch.

29. The sgRNA of item 28, wherein the hybrid of the first stem portion comprises or consists of the sequence 5'-GUCUUUGUA-3' hybridized to the sequence 5'-UACAAAGAU-3'.

30. The sgRNA of any one of items 1 to 29, further comprising one or more additional hairpin-forming segments located 3' to the first hairpin-forming segment.

31. The sgRNA of item 30, further comprising one or more linker segments located between the first hairpin-forming segment and additional hairpin-forming segments, and/or between the additional hairpin-forming segments.

32. A nucleic acid comprising a nucleotide sequence encoding the sgRNA of any one of items 1 to 31.

33. A vector comprising the nucleic acid of item 32.

34. The vector of item 33, further comprising a nucleotide sequence encoding a CRISPR nuclease.

35. The vector of item 34, wherein the CRISPR nuclease is a Cas9 enzyme.

36. The vector of item 34 or 35, wherein the CRISPR nuclease is derived from non-pathogenic bacteria.

37. The vector of any one of items 34 to 36, wherein the CRISPR nuclease is a *Streptococcus thermophilus* Cas9 nuclease.

38. The vector of any one of items 34 to 37, wherein the CRISPR nuclease is a type II Cas9 nuclease.

39. The vector of any one of items 34 to 38, wherein the CRISPR nuclease is a *Streptococcus thermophilus* type II-A CRISPR1-Cas9 (St1Cas9).

40. The vector of any one of items 34 to 39, wherein the CRISPR nuclease further comprises one or more nuclear localization signal (NLS) and the vector further comprises one or more nucleotide sequences encoding the one or more NLSs.

41. The vector of item 40, wherein the CRISPR nuclease comprises a first NLS at its amino terminal end and a second NLS at its carboxy terminal end, and the vector comprises NLS-encoding nucleotide sequences flanking the CRISPR nuclease-encoding nucleotide sequence.

42. The vector of item 33, further comprising a promoter operably-linked to the nucleotide sequence encoding the sgRNA.

43. The vector of any one of items 34 to 41, further comprising one or more promoters operably-linked to the nucleotide sequence encoding the sgRNA and or the nucleotide sequence encoding the CRISPR nuclease.

44. The vector of item 43, wherein the nucleotide sequence encoding the sgRNA and or the nucleotide sequence encoding the CRISPR nuclease are both operably linked to a single promoter.

45. The vector of item 43, wherein the nucleotide sequence encoding the sgRNA is operably linked to a first promoter and the nucleotide sequence encoding the CRISPR nuclease is operably linked to a second promoter, wherein the first and second promoters may be the same or different.

46. The vector of item 45, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in the same orientation within the vector.

47. The vector of item 45, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in opposite orientations within the vector.

48. The vector of any one of items 33 to 47, wherein the vector is a viral vector.

49. The vector of item 48, wherein the vector is an adeno-associated virus (AAV) vector.

50. A host cell comprising the nucleic acid of item 32 or the vector of any one of items 33 to 49.

51. A composition comprising the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, or the host cell of item 50.

52. The composition of item 51, further comprising a pharmaceutically acceptable carrier.

53. A system comprising the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the host cell of item 50, and/or the composition of claim 51 or 52.

54. A system comprising the vector of item 33 and a further vector comprising a nucleotide sequence encoding a CRISPR nuclease.

55. The system of item 54, wherein the CRISPR nuclease is as defined in any one of items 35 to 41.

56. The system of item 54 or 55, wherein the vector of item 33 further comprises a promoter operably-linked to the nucleotide sequence encoding the sgRNA and further vector further comprises a promoter operably-linked to the nucleotide sequence encoding the CRISPR nuclease.

57. A method of modifying a target polynucleotide in a cell, comprising contacting the cell with the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56.

58. The method of item 57, which is an *in vitro* method.

59. The method of item 57, which is an *in vivo* method and the cell is in a subject.

60. Use of the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56, for modifying a target polynucleotide in a cell.

61. Use of the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56, for the preparation of a medicament for modifying a target polynucleotide in a cell.

62. The sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56, for use in modifying a target polynucleotide in a cell.

63. A method of preventing or treating a condition associated with a target polynucleotide in a subject in need thereof, comprising administering to the subject an effective amount the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56.

64. The method of item 63, wherein the condition is a metabolic condition.

65. The method of item 63 or 64, wherein the condition is a hepatic condition.

66. Use of the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56, for preventing or treating a condition associated with a target polynucleotide in a subject.

67. Use of the sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56, for the preparation of a medicament for preventing or treating a condition associated with a target polynucleotide in a subject.

68. The use of item 66 or 67, wherein the condition is a metabolic condition.

69. The use of any one of items 66 to 68, wherein the condition is a hepatic condition.

70. The sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the composition of item 51 or 52 and/or the system of any one of items 53 to 56, for use in preventing or treating a condition associated with a target polynucleotide in a subject.

71. The sgRNA, nucleic acid, vector, composition and/or system for use of item 70, wherein the condition is a metabolic condition.

72. The sgRNA, nucleic acid, vector, composition and/or system for use of item 70 or 71, wherein the condition is a hepatic condition.

73. The sgRNA of any one of items 1 to 31, the nucleic acid of item 32, the vector of any one of items 33 to 49, the host cell of item 50, the composition of item 51 or 52 and/or the system of any one of items 53 to 56, for use as a medicament.

74. An isolated CRISPR nuclease polypeptide comprising a first domain and a second domain C-terminal to the first domain, wherein the first domain comprises a guide RNA-binding domain and a nuclease domain, and the second domain comprises a WED domain and a PAM-interacting domain, wherein the first and second domains are derived from different bacterial strains.

75. The isolated polypeptide of item 74, wherein the first and second domains are derived from non-pathogenic bacteria.

76. The isolated polypeptide of item 74 or 75, wherein the first and second domains are derived from different bacterial species.

77. The isolated polypeptide of item 74 or 75, wherein the first and second domains are derived from different strains of the same bacterial species.

78. The isolated polypeptide of item 77, wherein the first and second domains are derived from different strains of *Streptococcus thermophilus*.

79. The isolated polypeptide of any one of items 74 to 78, wherein the CRISPR nuclease is a type II Cas9 nuclease.

80. The isolated polypeptide of any one of items 74 to 79, wherein the CRISPR nuclease is a *Streptococcus thermophilus* type II-A CRISPR1-Cas9 (St1Cas9).

81. The isolated polypeptide of any one of items 74 to 80, further comprising one or more nuclear localization signal (NLS).

82. The isolated polypeptide of item 81, comprising a first NLS N-terminal to the first domain and a second NLS C-terminal to the second domain.

83. The isolated polypeptide of any one of items 74 to 82, further comprising a cytidine deaminase domain or an adenosine deaminase domain.

84. The isolated polypeptide of item 83, comprising a cytidine deaminase domain.

85. The isolated polypeptide of item 84, wherein the cytidine deaminase is an APOBEC cytidine deaminase.

86. The isolated polypeptide of item 84, wherein the cytidine deaminase domain comprises the amino acid sequence of SEQ ID NO: 50, or a functional fragment thereof, or a functional variant thereof.

87. The isolated polypeptide of item 84 or 85, further comprising a uracil DNA glycosylase inhibitor (UGI) domain.

88. The isolated polypeptide of item 87, wherein the UGI domain comprises the amino acid sequence of SEQ ID NO: 51, or a functional fragment thereof, or a functional variant thereof.

89. The isolated polypeptide of any one of items 74 to 88, wherein the first domain is derived from *Streptococcus thermophilus* LMD-9, LMG18311, CNRZ1066 or TH1477.

90. The isolated polypeptide of any one of items 74 to 89, wherein the second domain is derived from *Streptococcus thermophilus* LMD-9, LMG18311, CNRZ1066 or TH1477.

91. The isolated polypeptide of any one of items 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* LMD-9 and the second domain is derived from *Streptococcus thermophilus* LMG18311, CNRZ1066 or TH1477.

92. The isolated polypeptide of any one of items 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* LMG18311 and the second domain is derived from *Streptococcus thermophilus* LMD-9, CNRZ1066 or TH1477.

93. The isolated polypeptide of any one of items 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* CNRZ1066 and the second domain is derived from *Streptococcus thermophilus* LMG18311, LMD-9 or TH1477.

94. The isolated polypeptide of any one of items 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* TH1477 and the second domain is derived from *Streptococcus thermophilus* LMG18311, CNRZ1066 or LMD-9.

95. The isolated polypeptide of any one of items 74 to 88, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 264, 265, 266, or 267, or a functional fragment of any thereof, or a functional variant of any thereof.

96. The isolated polypeptide of any one of items 74 to 88, wherein the second domain comprises the amino acid sequence of SEQ ID NO: 260, 261, 262, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

97. The isolated polypeptide of any one of items 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 264, or a functional fragment of any thereof, or a functional variant of any thereof, and the second domain comprises the amino acid sequence of SEQ ID NO: 261, 262, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

98. The isolated polypeptide of any one of items 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 265, or a functional fragment of any thereof, or a functional variant of any thereof, and the second domain comprises the amino acid sequence of SEQ ID NO: 260, 262, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

99. The isolated polypeptide of any one of items 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 266, or a functional fragment of any thereof, or a functional variant of any thereof, and the second domain comprises the amino acid sequence of SEQ ID NO: 260, 261, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

100. The isolated polypeptide of any one of items 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 267, or a functional fragment of any thereof, or a functional variant of any thereof, and the second domain comprises the amino acid sequence of SEQ ID NO: 260, 261, or 262, or a functional fragment of any thereof, or a functional variant of any thereof.

101. The isolated polypeptide of any one of items 74 to 100, the first domain is connected to the second domains via a linker region.

102. The isolated polypeptide of any one of items 74 to 101, wherein the polypeptide is capable of binding a PAM that is different from the PAM bound by a CRISPR nuclease from which the first domain is derived.

103. The isolated polypeptide of any one of items 74 to 102, wherein the polypeptide binds a PAM comprising the sequence NNAGAA, NNGGAA, NNACAA, NNGCAA, NNGAAA or NNAAAA.

104. A nucleic acid comprising a nucleotide sequence encoding the isolated polypeptide of any one of items 74 to 103.

105. A vector comprising the nucleic acid of item 104.

106. The vector of item 105, further comprising a nucleotide sequence encoding an sgRNA.

107. The vector of item 106, wherein the sgRNA is the sgRNA of any one of items 1 to 31.

108. The vector of any one of items 105 to 107, further comprising one or more promoters operably-linked to the nucleotide sequence encoding the polypeptide and/or the nucleotide sequence encoding the sgRNA.

109. The vector of item 108, wherein the nucleotide sequence encoding the polypeptide and the nucleotide sequence encoding the sgRNA are both operably linked to a single promoter.

110. The vector of item 108, wherein the nucleotide sequence encoding the sgRNA is operably linked to a first promoter and the nucleotide sequence encoding the polypeptide is operably linked to a second promoter, wherein the first and second promoters may be the same or different.

111. The vector of item 110, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in the same orientation within the vector.

112. The vector of item 110, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in opposite orientations within the vector.

113. The vector of any one of items 105 to 112, wherein the vector is a viral vector.

114. The vector of item 113, wherein the vector is an adeno-associated virus (AAV) vector.

115. A host cell comprising the nucleic acid of item 104 or the vector of any one of items 105-113.

116. A composition comprising the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, or the host cell of item 115.

117. The composition of item 116, further comprising a pharmaceutically or biologically acceptable carrier.

118. A system comprising the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117.

119. A system comprising the vector of item 105 and a further vector comprising a nucleotide sequence encoding an sgRNA.

120. A method of modifying a target polynucleotide in a cell, comprising contacting the cell with the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119.

121. The method of item 120, which is an *in vitro* method.

122. The method of item 120, which is an *in vivo* method and the cell is in a subject.

123. Use of the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119, for modifying a target polynucleotide in a cell.

124. Use of the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119, for the preparation of a medicament for modifying a target polynucleotide in a cell.

125. The polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119, for use in modifying a target polynucleotide in a cell.

126. A method of preventing or treating a condition associated with a target polynucleotide in a subject in need thereof, comprising administering to the subject an effective amount the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119.

127. The method of item 126, wherein the condition is a metabolic condition.

128. The method of item 126 or 127, wherein the condition is a hepatic condition.

129. Use of the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119, for preventing or treating a condition associated with a target polynucleotide in a subject.

130. Use of the polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119, for the preparation of a medicament for preventing or treating a condition associated with a target polynucleotide in a subject.

131. The use of item 129 or 130, wherein the condition is a metabolic condition.

132. The use of any one of items 129 to 131, wherein the condition is a hepatic condition.

133. The polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119, for use in preventing or treating a condition associated with a target polynucleotide in a subject.

134. The sgRNA, nucleic acid, vector, composition and/or system for use of item 133, wherein the condition is a metabolic condition.

135. The sgRNA, nucleic acid, vector, composition and/or system for use of item 133 or 134, wherein the condition is a hepatic condition.

136. The polypeptide according to any one of items 74 to 103, the nucleic acid of item 104, the vector of any one of items 105 to 112, the host cell of item 115, and/or the composition of item 116 or 117, and/or the system of item 118 or 119, for use as a medicament.

Other objects, advantages and features of the present disclosure will become more apparent upon reading of the following non-restrictive description of specific embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF DRAWINGS

In the appended drawings:

FIG. 1: Engineered CRISPR1-StCas9 system drives robust gene editing in human cells. (**a**) Schematic representation of St1Cas9 from the LMD-9 strain flanked by nuclear localization signals (NLS). (**b**) Nucleotide sequence, predicted secondary structure, and functional modules of St1Cas9 sgRNA (v1; SEQ ID NO: 1); crRNA (up to position 34; left side of lower stem, bulge and upper stem), loop (positions 35-38; connecting left and right sides of upper stem), tracrRNA (position 39 and onward; right side of lower stem, bulge and upper stem, as well as stemloop 1, linker and stemloop 2), mutated nucleotides (positions 23 and 34). (**c**) St1Cas9 target sites (*FANCF*: sense, SEQ ID NO: 2; antisense, SEQ ID NO: 3; *EMX1*: sense, SEQ ID NO: 66; antisense, SEQ ID NO: 67; *RUNX1*: sense, SEQ ID NO: 68; antisense, SEQ ID NO: 69) and PAM sequences in *FANCF*, *EMX1*, and *RUNX1*. (**d**) K562 cells stably expressing St1Cas9 were transfected with indicated sgRNA expression vectors at increasing doses and the Surveyor and TIDE assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control.

FIG. 2: Screening for active sgRNAs targeting genes affecting liver function in mouse cells. (**a**) Surveyor assays to determine St1Cas9 activity programmed with various sgRNAs targeting *Pck1*. Neuro-2a cells were transiently transfected with a single vector (0.5µg) driving the expression of St1Cas9 and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. (**b**) Same as in (**a**) but targeting *Pcsk9*. (**c**) Same as in (**a**) but targeting *Hpd*.

FIG. 3: *In vivo* metabolic pathway rewiring via rAAV8-mediated delivery of holo-St1Cas9. (**a**) The tyrosine degradation pathway and associated genetic disorders. (**b**) Experimental design for *in vivo* editing. Neonatal (2 days old) *Fah*$^{-/-}$ mice were injected with rAAV8-St1Cas9 or saline into the retro-orbital sinus, weaned at 21 days, and NTBC was removed at 30 days of age. Mice were assayed for phenotypic and metabolic correction and for gene disruption efficacy. Mice off NTBC were killed when they lost 20% of their body weight. (**c**) Schematic representations of the rAAV vector and St1Cas9 target site (G5) within exon 13 of *Hpd*. Target sequence (sense, SEQ ID NO: 4; antisense, SEQ ID NO: 5; amino acid, SEQ ID NO: 6), PAM and position of the I335M mutation (last amino acid shown, i.e. I of LLQI) causing type III tyrosinemia in humans are shown. Also annotated is the human thyroxine binding globulin (TBG) promoter, bovine growth hormone polyadenylation sequence (BGHpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. (**d**) Neonatal *Fah*$^{-/-}$ mice were injected into the retro-orbital sinus with either 5E10, 1E11, 2E11 or 4E11 vector genomes (vg) of rAAV8-St1Cas9 targeting *Hpd* exon 13 (G5) and killed 28 days following injection.

Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (**e**) Survival analysis following NTBC removal in mice treated as described in (**b**). Number of mice per group (n) and rAAV doses (vg) is indicated. (**f**) Same as in (**e**) but body weight was measured daily. Solid lines designate the mean and error bars are represented by shaded areas and denote s.e.m. (**g**) Same as in (**f**) but glycemia was monitored in non-fasted mice. (**h**) Same as in (**e**) but succinylacetone levels in urine were determined 15 days following NTBC removal. Samples were collected from the indicated treatment groups over a 24 hours period using metabolic cages.

FIG. 4: Alternative rAAV-St1Cas9 vector architectures can further improve potency. (**a**) Schematic representations of the second-generation rAAV-St1Cas9 (v2) vector of similar size to the parent AAV genome (~4.7kb). Annotated is the human thyroxine binding globulin (TBG) promoter, synthetic polyadenylation sequence (SpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. Neonatal (2 days old) *Fah*$^{-/-}$ mice were injected with 2E11 vg rAAV8-St1Cas9 v2 targeting *Hpd* exon 13 (G5) or saline into the retro-orbital sinus and killed 13 days post injection. Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (**b**) Same as in (**a**) but the TBG promoter was swapped for the composite liver-specific LP1b promoter to generate rAAV8-St1Cas9 v3.

FIG. 5: Engineered CRISPR1-StCas9 system drives robust gene editing in human cells. (**a**) Schematic representation of the targeted integration of tagged St1Cas9 and SaCas9 to the *AAVS1* safe harbor locus. The donor construct and the locus following cDNA addition are displayed. The first two exons of the *PPP1R12C* gene are shown as open boxes. Also annotated are the locations of the splice acceptor site (SA), 2A self-cleaving peptide sequence (2A), puromycin resistance gene (Puro), polyadenylation sequence (pA), human phosphoglycerate kinase 1 promoter (h*PGK1*), nuclear localization signals (NLS), and 3xFLAG-2xSTREP tandem affinity tag (Tag), homology arms left and right (HA-L, HA-R) are respectively 800 and 840 bp. (**b**) Western blots showing Cas9-tag protein expression in K562 clones and cells expressing only the tag (Mock). The FLAG M2 antibody was used to detect Cas9 and the tubulin antibody was used as a loading control. (**c**) Alignment of previously described sgRNA sequences (SEQ ID NOs: 7-11) for St1Cas9. (**d**) K562 cells were transiently transfected with an St1Cas9 expression vector (0.5µg) in addition to the indicated sgRNA expression plasmids (0.8µg). Surveyor and TIDE assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control.

FIG. 6: *In vivo* metabolic pathway rewiring via rAAV8-mediated delivery of holo-St1Cas9. (a) Schematic representations of the rAAV vector and St1Cas9 target site (G2) within exon 8 of *Hpd*. Target sequence (sense, SEQ ID NO: 12; antisense, SEQ ID NO: 13; amino acid, SEQ ID NO: 14), PAM and position of the Y160C mutation causing type III tyrosinemia in humans are shown. Note that while this region of the protein is well conserved between human and mouse, a phenylalanine is found at this position in mouse, rat, pig, and *C. elegans*. Also annotated is the human thyroxine binding globulin (TBG) promoter, bovine growth hormone polyadenylation sequence (BGHpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. (b) Neonatal *Fah-/-* mice were injected into the retro-

orbital sinus with 1E11 vector genomes (vg) of rAAV8-St1Cas9 targeting *Hpd* exon 8 (G2) and killed 28 days following injection. Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (c) Body weight was measured daily following NTBC removal in mice treated as in Figure 3. Number of mice per group (n) is indicated. Dots designate the mean and error bars denote s.e.m. (d) Same as in (c) but glycemia was monitored in non-fasted mice.

FIG. 7: sgRNAs for SaCas9 and St1Cas9 are not functionally interchangeable. (**a**) A stable K562 cell line constitutively expressing SaCas9 was transfected with expression vectors (0.25µg) for its cognate sgRNA or the St1Cas9 sgRNA. The Surveyor assay was performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. (**b**) Same as in (**a**) but in a K562 cell line constitutively expressing St1Cas9. These data indicate that an sgRNA programmed to specify cleavage by St1Cas9 at one site cannot recruit and induce cutting by SaCas9 at this same site and vice-versa.

FIG. 8: St1Cas9 LMD-9 is functional at both NNAGAA and NNGGAA PAM sequences in human cells. (**a**) Surveyor assays to determine St1Cas9 activity programmed with various sgRNAs targeting *FANCF*, *EMX1*, *VEGFA*, and *RUNX1* (**Tables 3-4**). K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1Cas9 and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. *Indicates a non-specific PCR amplification product that generates a signal in the Surveyor assay. This signal was subtracted from all quantifications. (**b**) sgRNAs specifying cleavage by St1Cas9 at PAMs with an NN linker were modified to test their functionality with an NNN linker (**Tables 3, 6**). Cleavage activity was determined as in (**a**).

FIG. 9: Amino acid sequence alignment of SaCas9 (SEQ ID NO: 15) with St1Cas9 from different strains (SEQ ID NOs: 16-18 corresponding to St1Cas9_LMD-9, St1Cas9_LMG18311 and St1Cas9_CNRZ1066, respectively). The secondary structure of SaCas95 (5CZZ) is displayed above the sequences, which are numbered according to the residues of SaCas9. Identical residues are highlighted in black. Alignment was performed with Clustal Omega[6] and ESPript[7].

FIG. 10: Amino acid sequence alignment of St1Cas9 from different strains (SEQ ID NOs: 16-19 corresponding to LMD_9, LMG_18311, CNRZ_1066, and TH1477, respectively). Identical residues are highlighted in black. The position of the WED and PAM-interacting domain (PI) are indicated by arrows. This region of the protein has diverged the most as compared to the N-terminal segment. In SaCas9, the PAM duplex is sandwiched between the WED and PI domains5. Alignment was performed with Clustal Omega[6] and ESPript[7].

FIG. 11: St1Cas9 vectors available from Addgene (guide sense and antisense sequences, SEQ ID NOs: 20-21).

FIG. 12: Total editing efficacy at *FANCF* as determined by the TIDE assay for **Fig. 1**.

FIG. 13: Total editing efficacy at *EMX1* as determined by the TIDE assay for **Fig. 1**.

FIG. 14: Total editing efficacy at *RUNX1* as determined by the TIDE assay for **Fig. 1**.

FIG. 15: Total editing efficacy as determined by the TIDE assay for **Fig. 5**.

FIG. 16: Nucleotide sequence of St1Cas9 of strain LMD-9 (SEQ ID NO: 22). SV40 NLS is uppercase and underlined SEQ ID NOs: 23-24); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 25); linker regions are in lowercase (linkers flanking St1Cas9, SEQ ID NOs: 26-27).

FIG. 17: Amino acid sequence of St1Cas9 of strain LMD-9 (SEQ ID NO: 28). SV40 NLS is uppercase and underlined (SEQ ID NO: 29); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 30); linker regions are in lowercase (linkers flanking St1Cas9, SEQ ID NOs: 31-32).

FIG. 18: St1Cas9 LMD-9 is functional at both NNAGAA and NNGGAA PAM sequences in human cells. (a) Results of surveyor assays to determine St1Cas9 activity programmed with various sgRNAs targeting various PAMs. K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1Cas9 LMD-9 and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels. An expression vector encoding EGFP (-) was used as a negative control.

FIG. 19: Rewiring St1Cas9 LMD-9 to target a distinct PAM sequence using variants. (a) Schematic representation of St1Cas9 from the LMD-9 (A), LMG18311 (B), and CNRZ 1066 (C) strains along with their predicted or experimentally determined PAMs. The hybrid protein (AB) containing the N-terminal of St1Cas9 LMD-9 and the C-terminal domain (WED + PI) of St1Cas9 LMG18311 and CNRZ 1066 are also represented. (b-g) Surveyor assays to determine the activity of St1Cas9 variants programmed with sgRNAs targeting different PAM in human cells. K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1Cas9's and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control.

FIG. 20: Rewiring St1Cas9 LMD-9 to target distinct PAM sequences using variants. (a) Predicted PAM specificity for various St1Cas9 variants based on SPAMALOT. (b) Results of TIDE assays to determine St1Cas9 TH1477 activity programmed with various sgRNAs targeting various PAMs. K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1Cas9 TH1477 and its sgRNA. TIDE assays were performed 3 days later to determine the frequency of indels. An expression vector encoding EGFP (-) was used as a negative control.

FIG. 21: Converting St1Cas9 LMD-9 to a cytosine base editor (CBE). St1BE4max programmed with sgRNAs targeting NNAGAA and NNGGAA PAMs in human cells. K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1BE4max LMD-9 and its sgRNA. Quantification of base editing from sanger sequencing reads was performed 3 days later using EditR™ software. Numbers in each box indicate the % of C to T conversions. Protospacer target sequence SEQ ID NOs: 182, 180 and 178 (RUNX1); 172, 171, 169 and 168 (FANCF); 157, 162, 161, 159 and 160 (EMX1); 185 (ATP1A1).

FIG. 22: Rewiring St1BE4max to target distinct PAM sequences using variants. (a-b) St1BE4max LMG 18311 and CNRZ 1066 were programmed with sgRNAs targeting NNGCAA and NNACAA PAMs, respectively. K562

cells were transiently transfected with a single vector (1µg) driving the expression of St1BE4max variants and its sgRNA. Quantification of base editing from sanger sequencing reads was performed 3 days later using the EditR™ software. Numbers in each box indicate the % of C to T conversions. Protospacer target sequence SEQ ID NOs: 237, 236 and 235 (RUNX1); 258 and 239 (Grin2B); 234 (FANCF); 241 and 240 (ATP1A1, panel a); 238 (AAVS1); 243 and 242 (EMX1); 245 (ATP1A1, panel b).

FIGs. 23a-23b: Nucleotide sequence of NLS-St1Cas9 LMD-9/LMG18311 Hybrid-NLS (SEQ ID NO: 33). SV40 NLS is uppercase and underlined SEQ ID NOs: 23-24); St1Cas9 hybrid sequence is in uppercase italic (SEQ ID NO: 34); linker regions are in lowercase (linkers flanking St1Cas9 hybrid, SEQ ID NOs: 26-27).

FIG. 24: Amino acid sequence of NLS-St1Cas9 LMD-9/LMG18311 Hybrid-NLS (SEQ ID NO: 35). SV40 NLS is uppercase and underlined (SEQ ID NO: 29); St1Cas9 hybrid sequence is in uppercase italic (SEQ ID NO: 36); linker regions are in lowercase (linkers flanking St1Cas9, SEQ ID NOs: 31-32).

FIGs. 25a-25b: Nucleotide sequence of NLS-St1Cas9 LMD-9/CNRZ1066 Hybrid-NLS (SEQ ID NO: 37). SV40 NLS is uppercase and underlined SEQ ID NOs: 23-24); St1Cas9 hybrid sequence is in uppercase italic (SEQ ID NO: 38); linker regions are in lowercase (linkers flanking St1Cas9 hybrid, SEQ ID NOs: 26-27).

FIG. 26: Amino acid sequence of NLS-St1Cas9 LMD-9/CNRZ1066 Hybrid-NLS (SEQ ID NO: 30). SV40 NLS is uppercase and underlined (SEQ ID NO: 29); St1Cas9 hybrid sequence is in uppercase italic (SEQ ID NO: 40); linker regions are in lowercase (linkers flanking St1Cas9, SEQ ID NOs: 31-32).

FIGs. 27a-27b: Nucleotide sequence of NLS-St1Cas9 LMD-9/TH1477 Hybrid-NLS (SEQ ID NO: 41). SV40 NLS is uppercase and underlined SEQ ID NOs: 23-24); St1Cas9 hybrid sequence is in uppercase italic (SEQ ID NO: 42); linker regions are in lowercase (linkers flanking St1Cas9 hybrid, SEQ ID NOs: 26-27).

FIG. 28: Amino acid sequence of NLS-St1Cas9 LMD-9/TH1477 Hybrid-NLS (SEQ ID NO: 43). SV40 NLS is uppercase and underlined (SEQ ID NO: 29); St1Cas9 hybrid sequence is in uppercase italic (SEQ ID NO: 44); linker regions are in lowercase (linkers flanking St1Cas9, SEQ ID NOs: 31-32).

FIGs. 29a-29b: Nucleotide sequence of NLS-rAPOBEC1-St1Cas9 LMD-9-NLS-2xUGI-NLS-3xHA (SEQ ID NO: 45). SV40 NLS is uppercase and underlined SEQ ID NO: 23); rAPOBEC1 sequence is uppercase bold (SEQ ID NO: 46); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 25); UGI sequence is in uppercase bold italic (SEQ ID NO: 47); 3xHA sequence is in uppercase bold italic underlined (SEQ ID NO: 48).

FIG. 30: Amino acid sequence of NLS-rAPOBEC1-St1Cas9 LMD-9-NLS-2xUGI-NLS-3xHA (SEQ ID NO: 49). SV40 NLS is uppercase and underlined SEQ ID NO: 23); rAPOBEC1 sequence is uppercase bold (SEQ ID NO: 50); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 30); UGI sequence is in uppercase bold italic (SEQ ID NO: 51); 3xHA sequence is in uppercase bold italic underlined (SEQ ID NO: 52).

FIGs. 31a-31b: Nucleotide sequence of NLS-rAPOBEC1-St1Cas9 LMD-9/LMG18311-NLS-2xUGI-NLS-3xHA (SEQ ID NO: 53). SV40 NLS is uppercase and underlined SEQ ID NO: 23); rAPOBEC1 sequence is uppercase bold (SEQ ID NO: 47); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 54); UGI sequence is in uppercase bold italic (SEQ ID NO: 48); 3xHA sequence is in uppercase bold italic underlined (SEQ ID NO: 49).

FIG 32: Amino acid sequence of NLS-rAPOBEC1-St1Cas9 LMD-9/LMG18311-NLS-2xUGI-NLS-3xHA (SEQ ID NO: 55). SV40 NLS is uppercase and underlined SEQ ID NO: 23); rAPOBEC1 sequence is uppercase bold

(SEQ ID NO: 51); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 56); UGI sequence is in uppercase bold italic (SEQ ID NO: 52); 3xHA sequence is in uppercase bold italic underlined (SEQ ID NO: 53).

FIGs. 33a-33b: Nucleotide sequence of NLS-rAPOBEC1-St1Cas9 LMD-9/CNRZ1066-NLS-2xUGI-NLS-3xHA (SEQ ID NO: 57). SV40 NLS is uppercase and underlined SEQ ID NO: 23); rAPOBEC1 sequence is uppercase bold (SEQ ID NO: 47); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 58); UGI sequence is in uppercase bold italic (SEQ ID NO: 48); 3xHA sequence is in uppercase bold italic underlined (SEQ ID NO: 49).

FIG. 34: Amino acid sequence of NLS-rAPOBEC1-St1Cas9 LMD-9/CNRZ1066-NLS-2xUGI-NLS-3xHA (SEQ ID NO: 59). SV40 NLS is uppercase and underlined SEQ ID NO: 23); rAPOBEC1 sequence is uppercase bold (SEQ ID NO: 51); St1Cas9 sequence is in uppercase italic (SEQ ID NO: 60); UGI sequence is in uppercase bold italic (SEQ ID NO: 52); 3xHA sequence is in uppercase bold italic underlined (SEQ ID NO: 53).

FIG. 35: Domain organization of St1Cas9 from *S. thermophilus* LMD-9. BH: bridge helix, CTD: C-terminal domain, PI: PAM-interacting domain, WED: wedge domain. (a) Schematic representation of St1Cas9 domains; (b) amino acid sequence alignment of the C-terminal regions (including WED and PAM-interacting domain (PI) – see FIG. 10) of St1Cas9 from different strains (SEQ ID NOs: 260-263 corresponding to C-terminal regions of LMD_9, LMG_18311, CNRZ_1066, and TH1477, respectively). Identical residues are highlighted in black.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Described herein are reagents and methods for genetic modification using a CRISPR-Cas9 system. For example, CRISPR-based genetic modification is shown herein both *in vitro* and *in vivo*.

In an aspect, described herein are modified sgRNA architectures for CRISPR-based genetic modification. Therefore, in an aspect, described herein is an sgRNA, e.g., for modification of a target polynucleotide in a cell, comprising:

(a) a guide segment comprising a guide sequence corresponding to a region of the target polynucleotide;

(b) a first hairpin-forming segment located 3' to the guide sequence, the first hairpin hairpin-forming segment being capable of forming a hairpin comprising a stem portion and a loop portion, wherein the stem portion does not comprise a sequence corresponding to an RNA polymerase III termination signal. RNA polymerase III terminates at a poly(T) stretch, of typically 5-6 nucleotides in length. A poly(T) stretch on the target would correspond to a poly(U) in the sgRNA. Thus in an embodiment, the stem portion does not comprise more than 4 consecutive uracil nucleotides (U's), in a further embodiment, the stem portion does not comprise more than 3 consecutive U's.

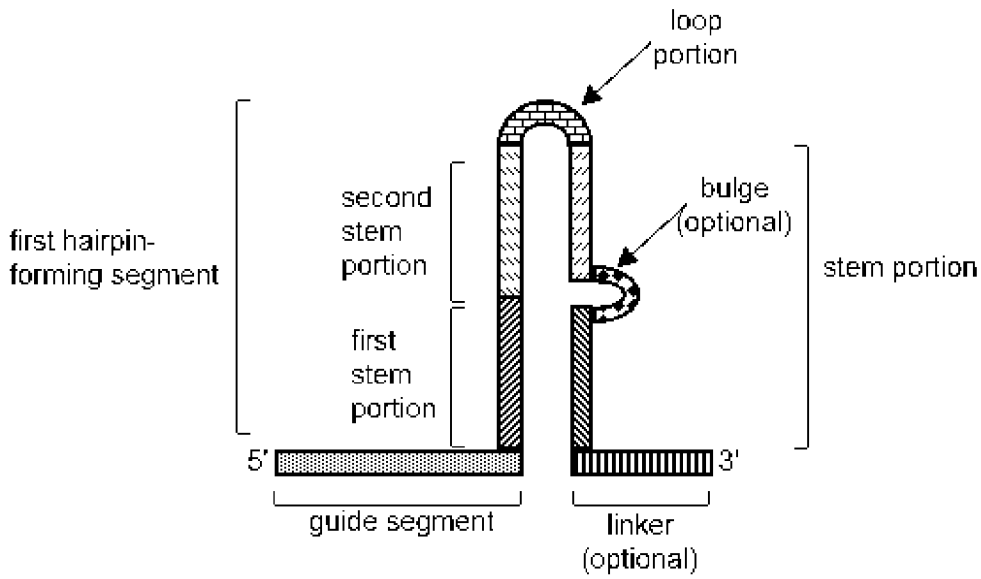
In an embodiment the cell is a eukaryotic cell, in a further embodiment, a mammalian cell, in a further embodiment, a human cell. In further embodiments, the cell is a fungal (e.g., yeast), plant or animal cell.

A hairpin (or stem-loop) forms when the phosphodiester backbone folds back onto itself to form a double-helical tract (the stem), leaving unpaired nucleotides to form a single-stranded "loop" region.

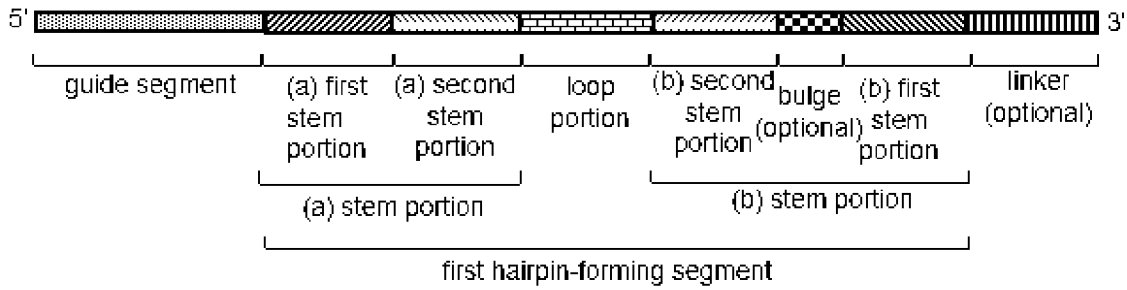
The stem may be subdivided into first and second stem portions (e.g. lower and upper stem portions, when considering a hairpin illustrated in an upright orientation).

The first hairpin may optionally comprise a bulge portion separating the first and second stem portions. Bulges and internal loops for when two double-helical tracts are separated on either one or both strands, due to one or more unpaired nucleotides.

In an embodiment, such an sgRNA can be illustrated schematically as follows, with the optional bulge and linker shown, when it has adopted a hairpin configuration:



The same sgRNA in a linear configuration appears as follows:



In the above schematic, (a) and (b) denote the two strands of the stem portion, created when the single strand folds back onto itself to create a two-strand hybrid or duplex structure. Thus the (a) and (b) portions are at least partially complementary to each other to enable formation of the stem portion.

In an embodiment, a predicted secondary structure of an sgRNA is shown in Fig. 1b, with the "guide" corresponding to the guide segment, the "lower" and "upper" stems corresponding to the first and second stem portions, respectively, the "GUAC" loop corresponding to the loop portion, and also showing the bulge. In embodiments, further secondary structures may be formed downstream (3' to the first hairpin-forming segment), as shown in Fig. 1b as "stemloop 1" and "stemloop 2".

In embodiments, the loop portion comprises or consists of a sequence of 3 to 6 nucleotides, in a further embodiment, 3 to 5 nucleotides, in a further embodiment, 4 nucleotides.

In embodiments, such a loop comprises or consists of the nucleotide sequence $N^1N^2N^3N^4$, wherein N^1, N^2, and N^3 are each independently A, C, G or U, and N^4 is C or G. In a further embodiment, N^1, N^3, and N^4 are each independently A, C, G or U, and N^2 is U, G or A. In a further embodiment, N^1 is G. In a further embodiment, N^2 is U. In a further embodiment, N^3 is A. In a further embodiment, N^4 is C. In an embodiment, such a loop comprises or consists of the sequence GUAC.

In embodiments, the second stem portion comprises or consists of a hybrid of 4 nucleotide pairs. In an embodiment, the fourth pair of the hybrid of the second stem portion, distal to the first stem portion, is a G-C pair. In a further embodiment, the hybrid of the second stem portion comprises or consists of the sequence 5'-UCUG-3' hybridized to the sequence 5'-CAGA-3'.

In an embodiment, the first stem portion comprises or consists of a hybrid of at least 5 nucleotide pairs. In a further embodiment, the first stem portion comprises or consists of a hybrid of not more than 12 nucleotide pairs. In further embodiments, the first stem portion comprises or consists of a hybrid of 6 to 10, 7 to 9, 5, 6, 7, 8, 9, 10, 11 or 12 nucleotide pairs. In an embodiment, the hybrid of the first stem portion comprises or consists of the sequence 5'-UCUUUGUA-3' hybridized to the sequence 5'-UACAAAGA-3'. In an embodiment, the hybrid of the first stem portion comprises or consists of the sequence 5'-GUCUUUGUA-3' hybridized to the sequence 5'-UACAAAGAU-3'.

In an embodiment the first stem portion does not comprise a mismatch. In an embodiment, the first stem portion comprises one or more mismatches, in a further embodiment, 1-2 mismatches, in a further embodiment, a single mismatch.

As noted above, in embodiments, the sgRNA further comprises one or more additional hairpin-forming segments located 3' to the first hairpin-forming segment. In embodiments, the sgRNA further comprises one or more linker segments located between the first hairpin-forming segment and additional hairpin-forming segments, and/or between the additional hairpin-forming segments.

Also described herein are nucleic acids comprising a nucleotide sequence encoding an sgRNA described herein.

Also described herein are vectors comprising a nucleic acid described herein. In an embodiment, the vector further comprises a nucleotide sequence encoding a CRISPR nuclease. In an alternative arrangement, two vectors may be used, one for expression of the sgRNA and the other for expression of the CRISPR nuclease, however a single vector for expression of both the sgRNA and CRISPR nuclease is preferred, particularly for *in vivo* applications.

In an embodiment, the CRISPR nuclease is derived from non-pathogenic bacteria. In an embodiment, the CRISPR nuclease is a Cas9 nuclease, in a further embodiment, a Cas9 nuclease from a non-pathogenic bacterium. In a further embodiment, the Cas9 nuclease is a *Streptococcus thermophilus* Cas9 nuclease. In a further embodiment, the Cas9 nuclease is a *Streptococcus thermophilus* type II-A CRISPR1-Cas9 (St1Cas9). The distinctive functional PAM sequences (NNAGAA and NNGGAA) of St1Cas9 increase the targeting flexibility and combinatorial potential of CRISPR-based genome editing tools.

Also described herein are engineered hybrid CRISPR nucleases combining gRNA-binding and nuclease domains from one source with a PAM-interacting domain from another source. This strategy allows for example the modification of PAM specificity of a CRISPR nuclease.

Therefore, in an aspect, there is further provided an isolated CRISPR nuclease polypeptide comprising a first domain and a second domain C-terminal to the first domain, wherein the first domain comprises a guide RNA-binding domain and a nuclease domain, and the second domain comprises a WED domain and a PAM-interacting domain.

In embodiments, the first and second domains the first and second domains are derived from different sources, i.e., they do not occur together in the same CRISPR nuclease in nature. In an embodiment, the first and second domains are from different bacterial strains, in a further embodiment, from different bacterial species, in a further embodiment, from different strains of the same bacterial species. In an embodiment, the first and second domains are derived from different strains of *Streptococcus thermophilus*.

The CRISPR nucleases described herein may also be used in a base editing approach, by using the CRISPR/Cas9 system to modify a cytidine (C) into a thymidine (T) in a target nucleic acid, or to modify an adenosine (A) into an inosine (I), which is read as a guanine (G), in a target sequence. In such an approach, an sgRNA may be designed and used in combination with a Cas9 nuclease (e.g. a Cas9 nickase) fused with a cytidine deaminase enzyme C into a T) or to modify an A into an I (read as G) in a target nucleic acid. Thus in embodiments a CRISPR nuclease or polypeptide described herein may further comprise a cytidine deaminase domain or an adenosine deaminase domain. In an embodiment, the cytidine deaminase is an APOBEC cytidine deaminase (e.g., comprising the amino acid sequence of SEQ ID NO: 50, or a functional fragment thereof, or a functional variant thereof). Further, enhanced C to T base editing may be achieved by co-expressing a uracil DNA glycosylase inhibitor (UGI). Thus in an embodiment, embodiments a CRISPR nuclease or polypeptide described herein may be used in conjunction with or fused to a UGI domain (e.g. comprising the amino acid sequence of SEQ ID NO: 51, or a functional fragment thereof, or a functional variant thereof).

In embodiments, an engineered hybrid CRISPR nuclease may comprise gRNA-binding and nuclease domains from *Streptococcus thermophilus* LMD-9, LMG18311, CNRZ1066 or TH1477. In further embodiments, an engineered hybrid CRISPR nuclease may comprise a PAM-interacting domain from *Streptococcus thermophilus* LMD-9, LMG18311, CNRZ1066 or TH1477. In embodiments, an engineered hybrid CRISPR nuclease may comprise:

- gRNA-binding and nuclease domains from *Streptococcus thermophilus* LMD-9 and a PAM-interacting domain derived from *Streptococcus thermophilus* LMG18311, CNRZ1066 or TH1477.
- gRNA-binding and nuclease domains from *Streptococcus thermophilus* LMG18311 and a PAM-interacting domain derived from *Streptococcus thermophilus* LMD-9, CNRZ1066 or TH1477.
- gRNA-binding and nuclease domains from *Streptococcus thermophilus* CNRZ1066 and a PAM-interacting domain derived from *Streptococcus thermophilus* LMG18311, LMD-9 or TH1477.
- gRNA-binding and nuclease domains from *Streptococcus thermophilus* TH1477 and a PAM-interacting domain derived from *Streptococcus thermophilus* LMG18311, CNRZ1066 or LMD-9.

In embodiments, the domain comprising the gRNA-binding and nuclease domains comprises the amino acid sequence of SEQ ID NO: 264, 265, 266, or 267, or a functional fragment of any thereof, or a functional variant of any thereof. In embodiments, the domain comprising the gRNA-binding and nuclease domains comprises an amino acid sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% or 98% identical to the amino acid sequence of SEQ ID NO: 264, 265, 266, or 267, which are embodiments of functional variants of SEQ ID NO: 264, 265, 266, and

267. In embodiments, the domain comprising the PAM-interacting domain comprises the amino acid sequence of SEQ ID NO: 260, 261, 262, or 263, or a functional fragment of any thereof, or a functional variant of any thereof. In embodiments, the domain comprising the PAM-interacting domain comprises an amino acid sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% or 98% identical to the amino acid sequence of SEQ ID NO: 260, 261, 262, or 263, which are embodiments of functional variants of SEQ ID NO: 260, 261, 262, and 263.

In embodiments, one or more linker regions (e.g., one or more amino acids) may be used to connect any of the domains described herein.

Also described herein are engineered hybrid CRISPR nucleases combining gRNA-binding and nuclease domains from one source with a PAM-interacting domain from another source. This strategy allows for example the modification of PAM specificity of a CRISPR nuclease. Thus the engineered polypeptide may be capable of binding a PAM that is different from the PAM bound by a CRISPR nuclease from which gRNA-binding and nuclease domains are derived. In embodiments, the engineered polypeptide binds a PAM comprising the sequence NNAGAA, NNGGAA, NNACAA, NNGCAA, NNGAAA or NNAAAA.

In embodiments, the domain comprising the PAM-interacting domain is derived from LMD-9 (e.g., SEQ ID NO: 260, or a functional fragment of any thereof, or a functional variant of any thereof) and is specific for NNAGAA and NNGGAA PAMs. In embodiments, the domain comprising the PAM-interacting domain is derived from CNRZ1066 (e.g., SEQ ID NO: 262, or a functional fragment of any thereof, or a functional variant of any thereof) and is specific for NNACAA PAMs. In embodiments, the domain comprising the PAM-interacting domain is derived from LMG18311 (e.g., SEQ ID NO: 261, or a functional fragment of any thereof, or a functional variant of any thereof) and is specific for NNGCAA PAMs. In embodiments, the domain comprising the PAM-interacting domain is derived from TH1477 (e.g., SEQ ID NO: 263, or a functional fragment of any thereof, or a functional variant of any thereof) and is specific for NNGAAA PAMs.

In embodiments, CRISPR nuclease (Cas or other nuclease/nickase recombinant protein described herein) preferably comprises at least one Nuclear Localization Signal (NLS) to target the protein into the cell nucleus, and the vector further comprises one or more nucleotide sequences encoding the one or more NLS's. Accordingly, as used herein the expression "nuclear localization signal" or "NLS" refers to an amino acid sequence, which 'tags' a protein for import into the cell nucleus by nuclear transport. Typically, this signal consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface. Different nuclear localized proteins may share the same NLS. An NLS has the opposite function of a nuclear export signal, which targets proteins out of the nucleus. Classical NLSs can be further classified as either monopartite or bipartite. The first NLS to be discovered was the sequence PKKKRKV (SEQ ID NO: 29) in the SV40 Large T-antigen (a monopartite NLS). The NLS of nucleoplasmin, KR[PAATKKAGQA]KKKK (SEQ ID NO: 61), is the prototype of the ubiquitous bipartite signal: two clusters of basic amino acids, separated by a spacer of about 10 amino acids. The Cas9 protein exemplified herein is a Cas9 nuclease comprising one or more, preferably two, NLS sequences.

There are many other types of NLS, which are qualified as "non-classical", such as the acidic M9 domain of hnRNP A1, the sequence KIPIK in yeast transcription repressor Matα2, the complex signals of U snRNPs as well as a recently identified class of NLSs known as PY-NLSs. Thus, any type of NLS (classical or non-classical) may be used

in accordance with the present disclosure as long as it targets the protein of interest into the nucleus of a target cell. In an embodiment, the NLS is derived from the simian virus 40 large T antigen. In an embodiment, the NLS of the recombinant protein of the present disclosure comprises or consists of the following amino acid sequence: SPKKKRKVEAS (SEQ ID NO: 62). In an embodiment the NLS comprises or consists of the sequence KKKRKV (SEQ ID NO: 63). In an embodiment, the NLS comprises or consists of the sequence SPKKKRKVEASPKKKRKV (SEQ ID NO: 64). In another embodiment, the NLS comprises or consists of the sequence KKKRK (SEQ ID NO: 65). In another embodiment, the NLS comprises comprises or consists of the sequence PKKKRKV (SEQ ID NO: 29).

In an embodiment, the CRISPR nuclease comprises a first NLS at its amino terminal end and a second NLS at its carboxy terminal end, and the vector comprises NLS-encoding nucleotide sequences flanking the CRISPR nuclease-encoding nucleotide sequence.

In embodiments, the vector further comprises one or more promoters operably-linked to the nucleotide sequence encoding the sgRNA and or the nucleotide sequence encoding the CRISPR nuclease. In an embodiment, the nucleotide sequence encoding the sgRNA and the nucleotide sequence encoding the CRISPR nuclease are both operably linked to a single promoter. In a further embodiment, the nucleotide sequence encoding the sgRNA is operably linked to a first promoter and the nucleotide sequence encoding the CRISPR nuclease is operably linked to a second promoter, wherein the first and second promoters may be the same or different. In the case where two promoters are used, (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease may be in the same orientation within the vector, in a further embodiment, they may be in opposite orientations within the vector.

In an embodiment, the vector is a viral vector, such as an adeno-associated virus (AAV) vector.

Also described herein are host cells comprising the nucleic acid(s) or vector(s) described herein.

Also described herein is a composition comprising an sgRNA, nucleic acid, vector, CRISPR nuclease and/or host cell described herein, which may optionally further comprise a biologically or pharmaceutically acceptable carrier.

Also described herein is a system or combination comprising an sgRNA, nucleic acid, vector, CRISPR nuclease host cell, and/or composition described herein

Also described herein are method of modifying a target polynucleotide in a cell, comprising contacting the cell with a sgRNA, nucleic acid, vector, CRISPR nuclease, host cell, composition and/or system or combination described herein.

In an embodiment, the method is an *in vitro* method. In a further embodiment, the method is an *in vivo* method and the cell is in a subject. In an embodiment, the method results in substantially no immune response in the subject.

Also described herein are a use of an sgRNA, nucleic acid, vector, CRISPR nuclease, host cell, composition and/or system or combination for modifying a target polynucleotide in a cell, or for the preparation of a composition or medicament for modifying a target polynucleotide in a cell. In an embodiment, the cell is in a subject and the use results in substantially no immune response in the subject.

The methods, uses and products described herein may be used to effect modifications in a target nucleic acid associated with a disease or condition, and therefore also provided herein are methods, uses and products for the prevention or treatment of a condition.

Therefore, also described herein is a method of treating a condition associated with a target polynucleotide in a subject in need thereof, comprising administering to the subject an effective amount an sgRNA, nucleic acid, vector, CRISPR nuclease, host cell, composition and/or system or combination described herein. In an embodiment, the method results in substantially no immune response in the subject.

Also described herein is a use of an sgRNA, nucleic acid, vector, CRISPR nuclease, host cell, composition and/or system or combination described herein, for use in preventing or treating a condition associated with a target polynucleotide in a subject, or for the preparation of a medicament for preventing or treating a condition associated with a target polynucleotide in a subject. In an embodiment, the use results in substantially no immune response in the subject.

Also described herein is an sgRNA, nucleic acid, vector, CRISPR nuclease, host cell, composition and/or system or combination described herein, for use as a medicament, e.g., for use in preventing or treating a condition described herein.

In embodiments, the condition is a metabolic condition, such as a condition affecting amino acid metabolism (e.g. tyrosine metabolism, e.g. a tyrosinemia). In an embodiment, the condition is a hepatic condition.

An “effective amount” includes a "therapeutically effective amount" and a "prophylactically effective amount". A "therapeutically effective amount" refers to an amount effective, at dosages and for periods of time necessary, to achieve the desired therapeutic result. A "prophylactically effective amount" refers to an amount effective, at dosages and for periods of time necessary, to achieve the desired prophylactic result, such as preventing or inhibiting the rate of onset or progression of a disease or condition. A prophylactically effective amount can be determined as described above for the therapeutically effective amount.

As used herein, the terms "subject" or "patient" are used interchangeably and are used to mean any animal, such as a mammal, including humans and non-human primates. In an embodiment, the above-mentioned subject is a mammal. In a further embodiment, the above-mentioned subject is a human.

DEFINITIONS

In order to provide clear and consistent understanding of the terms in the instant application, the following definitions are provided.

The use of the terms "a" and "an" and "the" and similar referents in the context of describing the subject matter disclosed herein (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context.

The use of the word “a” or “an” when used in conjunction with the term “comprising” in the claims and/or the specification may mean “one”, but it is also consistent with the meaning of “one or more”, “at least one”, and “one or more than one”. Similarly, the word “another” may mean at least a second or more.

As used in this specification and claim(s), the words “comprising” (and any form of comprising, such as “comprise” and “comprises”), “having” (and any form of having, such as “have” and “has”), “including” (and any form of including, such as “includes” and “include”) or “containing” (and any form of containing, such as “contains” and

"contain") are inclusive or open-ended and do not exclude additional, un-recited elements or method steps and are used interchangeably with the phrase "including but not limited to".

Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All subsets of values within the ranges are also incorporated into the specification as if they were individually recited herein. For example, for the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 18-20, the numbers 18, 19 and 20 are explicitly contemplated, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. Further, in embodiments various steps may be repeated, to for example increase recovery and purification.

The use of any and all examples, or exemplary language (*e.g.*, "such as") provided herein, is intended merely to better illustrate the disclosure and does not pose a limitation on the scope of the disclosure unless otherwise claimed.

Any and all combinations and sub-combinations of the embodiments and features disclosed herein are encompassed by the present disclosure.

Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. For example, any nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those that are well known and commonly used in the art. The meaning and scope of the terms should be clear; in the event however of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

Practice of the methods, as well as preparation and use of the products and compositions disclosed herein employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. See, for example, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P. M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P. B. Becker, ed.) Humana Press, Totowa, 1999.

The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect

to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (e.g., phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; i.e., an analogue of A will base-pair with T.

The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of corresponding naturally-occurring amino acids.

As used herein, the term “non-conservative mutation” or “non-conservative substitution” in the context of polypeptides refers to a mutation in a polypeptide that changes an amino acid to a different amino acid with different biochemical properties (i.e., charge, hydrophobicity and/or size). Although there are many ways to classify amino acids, they are often sorted into six main groups on the basis of their structure and the general chemical characteristics of their R groups. (i) Aliphatic (Glycine, Alanine, Valine, Leucine, Isoleucine); (ii) Hydroxyl or Sulfur/Selenium-containing (also known as polar amino acids) (Serine, Cysteine, Selenocysteine, Threonine, Methionine); (iii) Cyclic (Proline); (iv) Aromatic (Phenylalanine, Tyrosine, Tryptophan); (v) Basic (Histidine, Lysine, Arginine) and (vi) Acidic and their Amide (Aspartate, Glutamate, Asparagine, Glutamine). Thus, a non-conservative substitution includes one that changes an amino acid of one group with another amino acid of another group (e.g., an aliphatic amino acid for a basic, a cyclic, an aromatic or a polar amino acid; a basic amino acid for an acidic amino acid, a negatively charged amino acid (aspartic acid or glutamic acid) for a positively charged amino acid (lysine, arginine or histidine) etc.

Conversely, a “conservative substitution” or “conservative mutations” in the context of polypeptides are mutations that change an amino acid to a different amino acid with similar biochemical properties (e.g. charge, hydrophobicity and size). For example, a leucine and isoleucine are both aliphatic, branched hydrophobes. Similarly, aspartic acid and glutamic acid are both small, negatively charged residues. Therefore, changing a leucine for an isoleucine (or vice versa) or changing an aspartic acid for a glutamic acid (or vice versa) are examples of conservative substitutions.

"Coding sequence" or "encoding nucleic acid" as used herein means the nucleic acids (RNA or DNA molecule) that comprise a nucleotide sequence which encodes a protein or sgRNA. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is administered. The coding sequence may be codon optimized, e.g. for use in eukaryotic, mammalian and/or human cells.

In embodiments, recombinant expression vectors of the disclosure can comprise a polynucleotide of the present disclosure in a form suitable for expression of the polynucleotide in a host cell, which means that the recombinant expression vector includes one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, “operably-linked” is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term “regulatory sequence” is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation

signal). Such regulatory sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in a certain host cell (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the present disclosure can be introduced into host cells to thereby produce sgRNAs, proteins or peptides, encoded by polynucleotides as described herein.

"Complement" or "complementary" as used herein refers to Watson-Crick (e.g., A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. "Complementarity" refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

Sequence similarity

"Homology" and “homologous” refers to sequence similarity between two peptides or two nucleic acid molecules. Homology can be determined by comparing each position in the aligned sequences. A degree of homology between nucleic acid or between amino acid sequences is a function of the number of identical or matching nucleotides or amino acids at positions shared by the sequences. As the term is used herein, a nucleic acid sequence is "substantially homologous" to another sequence if the two sequences are substantially identical and the functional activity of the sequences is conserved (as used herein, the term “homologous” does not infer evolutionary relatedness, but rather refers to substantial sequence identity, and thus is interchangeable with the terms “identity”/“identical”). Two nucleic acid sequences are considered substantially identical if, when optimally aligned (with gaps permitted), they share at least about 50% sequence similarity or identity, or if the sequences share defined functional motifs. In alternative embodiments, sequence similarity in optimally aligned substantially identical sequences may be at least 60%, 70%, 75%, 80%, 85%, 90% or 95%. For the sake of brevity, the units (e.g., 66, 67…81, 82,…91, 92%….) have not systematically been recited but are considered, nevertheless, within the scope of the present disclosure.

Substantially complementary nucleic acids are nucleic acids in which the complement of one molecule is substantially identical to the other molecule. Two nucleic acid or protein sequences are considered substantially identical if, when optimally aligned, they share at least about 70% sequence identity. In alternative embodiments, sequence identity may for example be at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 98% or at least 99%. Optimal alignment of sequences for comparisons of identity may be conducted using a variety of algorithms, such as the local homology algorithm of Smith and Waterman, 1981, Adv. Appl. Math 2: 482, the homology alignment algorithm of Needleman and Wunsch, 1970, J. Mol. Biol. 48:443, the search for similarity method of Pearson and Lipman (Pearson and Lipman 1988), and the computerized implementations of these algorithms (such as GAP, BESTFIT, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, WI, U.S.A.). Sequence identity may also be determined using the BLAST algorithm, described in Altschul et al. (Altschul et al. 1990) 1990 (using the published default settings). Software for performing

BLAST analysis may be available through the National Center for Biotechnology Information (through the internet at http://www.ncbi.nlm.nih.gov/). The BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold. Initial neighborhood word hits act as seeds for initiating searches to find longer HSPs. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction is halted when the following parameters are met: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. One measure of the statistical similarity between two sequences using the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. In alternative embodiments of the disclosure, nucleotide or amino acid sequences are considered substantially identical if the smallest sum probability in a comparison of the test sequences is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

An alternative indication that two nucleic acid sequences are substantially complementary is that the two sequences hybridize to each other under moderately stringent, or preferably stringent, conditions. Hybridization to filter-bound sequences under moderately stringent conditions may, for example, be performed in 0.5 M NaHPO4, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.2 x SSC/0.1% SDS at 42°C (Ausubel 2010). Alternatively, hybridization to filter-bound sequences under stringent conditions may, for example, be performed in 0.5 M NaHPO4, 7% SDS, 1 mM EDTA at 65°C, and washing in 0.1 x SSC/0.1% SDS at 68°C (Ausubel 2010). Hybridization conditions may be modified in accordance with known methods depending on the sequence of interest (Tijssen 1993). Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point for the specific sequence at a defined ionic strength and pH.

"Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid or between a sgRNA and a target polynucleotide or between a sgRNA and a CRISPR nuclease (e.g., Cas9, Cpf1). Not all components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is sequence-specific. “Affinity” refers to the strength of binding: increased binding affinity being correlated with a lower Kd.

A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, etc.) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

As used herein, “a nuclease-based modification” refers to a modification in a polynucleotide e.g., an endogenous gene locus or genomic sequence) which involves the introduction of a cut (e.g., a double-stranded break

in the polynucleotide) which ultimately will trigger a repair mechanism by the cell involving (Non-homologous-end-joining) NHEJ or homologous recombination (HDR). The nuclease-based modification is made by site specific nucleases targeting the polynucleotide of interest (i.e., an endogenous gene locus or genomic sequence). Site-specific nucleases (engineered) are well known and include (but are not limited to) Zinc finger nucleases, meganucleases, Mega-Tals, CRISPR nucleases, TALENs, etc.

"Recombination" refers to a process of exchange of genetic information between two polynucleotides. For the purposes of this disclosure, "homologous recombination" (HR) refers to the specialized form of such exchange that takes place, for example, during repair of double-strand breaks in cells via homology-directed repair (HDR) mechanisms. This process requires nucleotide sequence homology, uses a "donor" or "patch" molecule as a template for repair of a "target" molecule (i.e., the one that experienced the double-strand break), and is variously known as "non-crossover gene conversion" or "short tract gene conversion," because it leads to the transfer of genetic information from the donor to the target. Without wishing to be bound by any particular theory, such transfer can involve mismatch correction of heteroduplex DNA that forms between the broken target and the donor, and/or "synthesis- dependent strand annealing," in which the donor is used to re-synthesize genetic information that will become part of the target, and/or related processes. Such specialized HR often results in an alteration of the sequence of the target molecule such that part or all of the sequence of the donor polynucleotide is incorporated into the target polynucleotide.

In the methods described herein, one or more targeted (site-specific) nucleases (e.g., sgRNA/CRISPR nuclease) create a double-stranded break in the target sequence (e.g., cellular chromatin) at a predetermined site. A "donor" polynucleotide, having homology to the nucleotide sequence in the region of the break, may be introduced into the cell if desired. The presence of the double-stranded break has been shown to facilitate integration of the donor sequence. The donor sequence may be physically integrated or, alternatively, the donor polynucleotide is used as a template for repair of the break via homologous recombination, resulting in the introduction of all or part of the nucleotide sequence as in the donor into the cellular chromatin. Thus, a first sequence in cellular chromatin can be altered and, in certain embodiments, can be converted into a sequence present in a donor polynucleotide. Thus, the use of the terms "replace" or "replacement" can be understood to represent replacement of one nucleotide sequence by another, (i.e., replacement of a sequence in the informational sense), and does not necessarily require physical or chemical replacement of one polynucleotide by another. In any of the methods described herein, additional sgRNA/CRISPR nucleases, pair zinc-finger, Meganucleases, Mega-Tals, and/or additional TALEN proteins can be used for additional double-stranded cleavage of additional target sites within the cell.

As used herein, the terms “donor” or “patch” nucleic acid are used interchangeably and refers to a nucleic acid that includes a fragment of the endogenous targeted gene of a cell (in some embodiments the entire targeted gene), but which includes desired modification(s) at specific nucleotides. The donor (patch) nucleic acid must be of sufficient size and similarity (e.g., in the right and left homology arms) to permit homologous recombination with the targeted gene. Preferably, the donor/patch nucleic acid is (or is flanked at the 5’ end and at the 3’ end by sequences) at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% identical to the endogenous targeted polynucleotide gene sequence. The patch nucleic acid may be provided for example as a ssODN, as a PCR product (amplicon) or within a vector. Preferably, the patch/donor nucleic acid will include modifications with respect to the

endogenous gene which i) precludes it from being cut by a sgRNA once integrated in the genome of a cell and/or which facilitate the detection of the introduction of the patch nucleic acid by homologous recombination.

As used herein, a “targeted gene”, “gene of interest” or “targeted polynucleotide” corresponds to the polynucleotide within a cell that will be modified by the introduction of the patch nucleic acid. It corresponds to an endogenous gene naturally present within a cell. The targeted gene may comprise one or more mutations associated with a risk of developing a disease or disorder which may be corrected by the introduction of the patch/donor nucleic acid (e.g., will be modified to correspond to the WT gene or to a form which is no longer associated with increased risk of developing a disease or condition). One or both alleles of a targeted gene may be corrected or modified within a cell in accordance with the present disclosure. Examples of target genes are described in Tables 3-6.

A “target polynucleotide” as used herein refers to any endogenous polynucleotide or nucleic acid present in the genome of a cell and encoding or not a known gene product. "Target gene" as used herein refers to any endogenous polynucleotide or nucleic acid present in the genome of a cell and encoding a known or putative gene product. The target gene or target polynucleotide further corresponds to the polynucleotide within a cell that will be modified by a nuclease of the present disclosure, alone or in combination with the introduction of one or more donor nucleic acid or patch nucleic acids. The target gene or target polynucleotide may be a mutated gene involved in a genetic disease.

"Promoter" as used herein means a synthetic or naturally-derived nucleic acid molecule which is capable of conferring, modulating or controlling (e.g., activating, enhancing and/or repressing) expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance or repress expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter, CMV IE promoter, U6 promoter, a liver-specific promoter (e.g., LP1b; combining the human apolipoprotein E/C-I gene locus control region (*ApoE*-HCR) and a modified human α1 antitrypsin promoter (h*AAT*) coupled to an SV40 intron), human thyroxine binding globulin (TBG) promoter, CMV promoter, CAG promoter, CBH promoter, UbiC promoter, Ef1a promoter, H1 promoter, and 7SK promoter, any of which may be used to express one or more sgRNAs and/or a CRISPR nuclease in a cell. Sequences for the LP1b and TBG promoters are provided in Table 8.

“Vector" as used herein means a nucleic acid sequence containing an origin of replication. A vector may be a viral vector, bacteriophage, bacterial artificial chromosome or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be a self- replicating extrachromosomal vector, and preferably, is a DNA plasmid. For example, the vector may comprise nucleic acid sequence(s) that/which encode(s) a sgRNA, a donor (or patch) nucleic

acid, and/or a CRISPR nuclease (e.g., Cas9 or Cpf1) of the present disclosure. A vector for expressing one or more sgRNA will comprise a “DNA” sequence of the sgRNA.

Nucleic acids encoding sgRNAs and CRISPR nucleases (e.g., Cas9) of the present disclosure may be delivered into cells using one or more various vectors such as viral vectors. Accordingly, preferably, the above-mentioned vector is a viral vector for introducing the gRNA and/or nuclease of the present disclosure in a target cell. Non-limiting examples of viral vectors include retrovirus, lentivirus, Herpes virus, adenovirus or Adeno Associated Virus, as well known in the art.

"Adeno-associated virus" or "AAV" as used interchangeably herein refers to a small virus belonging to the genus Dependovirus of the Parvoviridae family that infects humans and some other primate species. AAV is not currently known to cause disease and consequently the virus causes a very mild immune response.

In embodiments, the AAV vector preferably targets one or more cell types. Accordingly, the AAV vector may have enhanced cardiac, skeletal muscle, neuronal, liver, and/or pancreatic tissue (Langerhans cells) tropism. The AAV vector may be capable of delivering and expressing the at least one gRNA and nuclease of the present disclosure in the cell of a mammal. For example, the AAV vector may be an AAV-SASTG vector (Piacentino *et al.* (2012) Human Gene Therapy 23:635-646). The AAV vector may deliver gRNAs and nucleases to neurons, skeletal and cardiac muscle, and/or pancreas (Langerhans cells) *in vivo*. The AAV vector may be based on one or more of several capsid types, including AAVl, AAV2, AAV5, AAV6, AAV8, and AAV9. The AAV vector may be based on AAV2 pseudotype with alternative muscle-tropic AAV capsids, such as AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5 and AAV/SASTG vectors that efficiently transduce skeletal muscle or cardiac muscle by systemic and local delivery. In an embodiment, the AAV vector is a AAV-DJ. In an embodiment, the AAV vector is a AAV-DJ8 vector. In an embodiment, the AAV vector is a AAV2-DJ8 vector. In an embodiment, the AAV vector is a AAV-PHP.B vector. In an embodiment, the AAV vector is a AAV-PHP.B, AAV-9 or AAV-DJ8 (PHP.B: PMID: 26829320, PMID: 27867348; AAV DJ-8: www.cellbiolabs.com/news/aav-helper-free-expression-systems-aav-dj-aav-dj8, http://www.cellbiolabs.com/aav-expression-and-packaging; www.cellbiolabs.com/scaav-dj8-helper-free-complete-expression-systems; and AAV9: PMID: 27637390, PMID: 16713360).

In yet another aspect, the present disclosure provides a cell (e.g., a host cell) comprising the above-mentioned nucleic acid and/or vector. In embodiments, the host cell may be prokaryotic (e.g. bacteria) or eukaryotic (e.g., fungal (yeast), mammalian, murine, human). The disclosure further provides a recombinant expression system, vectors and host cells, such as those described above, for the expression/production of a recombinant protein, using for example culture media, production, isolation and purification methods well known in the art.

In another aspect, the present disclosure provides a composition (e.g., a pharmaceutical composition) comprising the above-mentioned gRNA, and/or CRISPR nuclease (e.g., Cas9), or nucleic acid(s) encoding same or vector(s) comprising such nucleic acid(s), or the above-mentioned host cells. In an embodiment, the composition further comprises one or more biologically or pharmaceutically acceptable carriers, excipients, and/or diluents.

As used herein, "pharmaceutically acceptable" (or "biologically acceptable") carriers, excipients, and/or diluents includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like that are physiologically compatible, and which can be used pharmaceutically

or in biological systems. Such materials are characterized by the absence of (or limited) toxic or adverse biological effects *in vivo*. It refers to those compounds, compositions, and/or dosage forms which are, within the scope of sound medical judgment, suitable for use in contact with the biological fluids and/or tissues and/or organs of a subject (e.g., human, animal) without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio.

When the excipient serves as a diluent, it can be a solid, semisolid, or liquid material, which acts as a vehicle, carrier or medium for the active ingredient. Thus, the compositions can be in the form of tablets, pills, powders, lozenges, sachets, cachets, elixirs, suspensions, emulsions, solutions, syrups, aerosols (as a solid or in a liquid medium), ointments containing for example up to 10% by weight of the active compound, soft and hard gelatin capsules, suppositories, sterile injectable solutions, and sterile packaged powders (see Remington: The Science and Practice of Pharmacy by Alfonso R. Gennaro, 2003, 21th edition, Mack Publishing Company). In embodiments, the carrier may be suitable for intra-neural, parenteral, intravenous, intraperitoneal, intramuscular, subcutaneous, sublingual or oral administration.

Some examples of suitable excipients include lactose, dextrose, sucrose, sorbitol, mannitol, starches, lecithin, phosphatidylcholine, gum acacia, calcium phosphate, alginates, tragacanth, gelatin, calcium silicate, microcrystalline cellulose, polyvinylpyrrolidone, cellulose, water, syrup and methyl cellulose. The formulations can additionally include: lubricating agents such as talc, magnesium stearate, and mineral oil; wetting agents; emulsifying and suspending agents; preserving agents such as methyl- and propylhydroxybenzoates; sweetening agents; and flavoring agents. The compositions of the disclosure can be formulated so as to provide quick sustained or delayed release of the active ingredient after administration to the patient by employing procedures known in the art.

Pharmaceutical compositions suitable for use in the disclosure include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose (*e.g.*, preventing, treating, ameliorating and/or inhibiting a disease or condition). The determination of an effective dose is well within the capability of those skilled in the art. For any compounds, the therapeutically effective dose can be estimated initially either in cell culture assays (*e.g.*, cell lines) or in animal models, usually mice, rabbits, dogs or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. An effective dose or amount refers to that amount of one or more active ingredient(s), which is sufficient for treating a specific disease or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, ED_{50} (the dose therapeutically effective in 50% of the population) and LD_{50} (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD_{50}/ED_{50}. Pharmaceutical compositions, which exhibit large therapeutic indices, are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration. The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety

or to maintain the desired effect. Factors, which may be taken into account, include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. In embodiments, dosages of an active ingredient of between about 0.01 and about 100 mg/kg body weight (in an embodiment, per day) may be used. In further embodiments, dosages of between about 0.5 and about 75 mg/kg body weight may be used. In further embodiments, dosages of between about 1 and about 50 mg/kg body weight may be used. In further embodiments, dosages of between about 10 and about 50 mg/kg body weight in further embodiments about 10, about 25 or about 50 mg/kg body weight, may be used.

The present disclosure further provides a kit or package comprising at least one container means having disposed therein at least one of the above-mentioned sgRNAs, nucleases, vectors, cells, systems, combinations or compositions. In an embodiment, the kit or package further comprises with instructions for use, such as for modification of a nucleotide sequence in a cell, or for the treatment of a condition associated with a target polynucleotide.

CRISPR system

CRISPR technology is a system for genome editing, e.g., for modification of a nucleic acid sequence, and may also be used for example to modify the expression of a specific gene.

This system stems from findings in bacterial and archaea which have developed adaptive immune defenses termed clustered regularly interspaced short palindromic repeats (CRISPR) systems, which use crRNAs and Cas proteins to degrade complementary sequences present in invading viral and plasmid DNA. The original CRISPR systems comprised a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA), which form a hybrid (which guides a CRISPR nuclease, e.g. a Cas9).

Engineered CRISPR systems use for example a synthetically reconstituted "guide RNA" ("sgRNA"), corresponding to a crRNA-tracrRNA fusion that obviates the need for RNase III and crRNA processing in general. The sgRNA comprises a “sgRNA guide sequence” or “sgRNA target sequence” and an RNA sequence (Cas recognition sequence)”, which is necessary for CRISPR nuclease (e.g., Cas9) binding to the targeted gene. The sgRNA guide sequence is the sequence that confers specificity. It hybridizes with (i.e., it is complementary to) the opposite strand of a target sequence (i.e., it corresponds to the RNA sequence of a DNA target sequence). Other CRISPR systems using different CRISPR nucleases have been developed and are known in the art (e.g., using the Cpf1 nuclease instead of a Cas9 nuclease).

Because the original Cas9 nuclease combined with a sgRNA may produce off-target mutagenesis, one may alternatively use in accordance with the present disclosure a pair of specifically designed sgRNAs in combination with a Cas9 nickase or in combination with a dCas9-FokI nuclease to cut both strands of DNA.

In embodiments, provided herein are CRISPR/nuclease-based engineered systems for use in modifying a target nucleic acid in cells. Introduction of DSBs can knockout a specific gene or allow modifying it by Homology Directed Repair (HDR), where one or more donor or patch nucleic acids comprising the desired modification(s) are provided to introduce the modification(s) by HDR. CRISPR/Cas9-induced DNA cleavage followed by Non-Homologous

End Joining (NHEJ) repair has been used to generate loss-of-function alleles in protein-coding genes or to delete a very large DNA fragment (20, 21). The CRISPR-based engineered systems of the present disclosure are designed to (i) target and cleave a gene of interest) to generate gene variants (e.g., creating insertion(s) and/or deletions, also referred to as INDELS).

Accordingly, in an aspect, the present disclosure involves the design and preparation of one or more sgRNAs for inducing a DSB (or two single stranded breaks (SSB) in the case of a nickase) in a target gene of interest. In embodiments, the present disclosure also involves the design and preparation of one or more sgRNAs for inducing a DSB (or two SSBs in the case of a nickase) in a target polynucleotide located at a different locus within the genome of target cells. The sgRNAs and the nuclease are then used together to introduce the desired modification(s) (i.e., gene-editing events) by NHEJ or HDR within the genome of one or more target cells. When the desired modification(s) include specific point mutation(s) or insertions/deletion(s), one or more donor or patch nucleic acids comprising the desired modification(s) are provided to introduce the modification(s) by HDR.

sgRNAs

In order to cut DNA at a specific site, CRISPR nucleases require the presence of a sgRNA and a protospacer adjacent motif (PAM) on the targeted gene. The PAM immediately follows (i.e., is adjacent to) the sgRNA target sequence in the targeted polynucleotide gene sequence. The PAM is located at the 3' end or 5' end of the sgRNA target sequence (depending on the CRISPR nuclease used) but is not included in the sgRNA guide sequence. For example, the PAM for Cas9 CRISPR nucleases is located at the 3' end of the sgRNA target sequence on the target gene while the PAM for Cpf1 nucleases is located at the 5' end of the sgRNA target sequence on the target gene. Different CRISPR nucleases also require a different PAM. Accordingly, selection of a specific polynucleotide sgRNA target sequence is generally based on the CRISPR nuclease used. The PAM for the *Streptococcus pyogenes* Cas9 CRISPR system is 5'-NRG-3', where R is either A or G, and characterizes the specificity of this system in human cells. The PAM of *S. aureus* Cas9 is NNGRR. The *S. pyogenes* Type II system naturally prefers to use an "NGG" sequence, where "N" can be any nucleotide, but also accepts other PAM sequences, such as "NAG" in engineered systems. Similarly, the Cas9 derived from *Neisseria meningitidis* (NmCas9) normally has a native PAM of NNNNGATT, but has activity across a variety of PAMs, including a highly degenerate NNNNGNNN PAM. The PAM for AsCpf1 or LbCpf1 CRISPR nuclease is TTTN. In an embodiment, the PAM for a Cas9 protein used in accordance with the present disclosure is a NGG trinucleotide-sequence (Cas9). In another embodiment, the PAM for a Cpf1 CRISPR nuclease used in accordance with the present disclosure is a TTTN nucleotide sequence. In a preferred embodiment, the St1Cas9 may be used, which corresponds to the PAM sequences NNAGAA and NNGGAA. In embodiments, different St1Cas9 PAM sequences may be used, for example, inferred consensus PAM sequences for St1Cas9 from strains CNRZ1066 and LMG13811 are NNACAA(W) and NNGCAA(A), respectively[24, 26]. Table 1 below provides a list of non-limiting examples of CRISPR/nuclease systems with their respective PAM sequences.

Table 1: Non-exhaustive list of CRISPR-nuclease systems from different species (see. Mohanraju, P. et al., PMID 27493190; Shmakov, S et al., PMID: 26593719; and Zetsche, B. et al., PMID: 26422227). Also included

are engineered variants recognizing alternative PAM sequences (see Kleinstiver, BP. et al., (Nature biotech 2015) PMID: 26524662 and Kleinstiver, BP. et al., (Nature 2015)).

CRISPR nuclease	PAM Sequence
Streptococcus pyogenes (SP); SpCas9	NGG + NAG
SpCas9 D1135E variant	NGG (reduced NAG binding)
SpCas9 VRER variant	NGCG
SpCas9 EQR variant	NGAG
SpCas9 VQR variant	NGAN or NGNG
Staphylococcus aureus (SA); SaCas9	NNGRRT or NNGRR(N)
SaCas9 KKH variant	NNNRRT
Neisseria meningitidis (NM)	NNNNGATT
Streptococcus thermophilus (ST1)	NNAGAA and NNGGAA
Treponema denticola (TD)	NAAAAC
AsCpf1(*Acidominococcus*)	TTTN
AsCpf1 S542R/K607R	TYCV
AsCpf1 S542R/K548V/N552R	TATV
LbCpf1 (*Lachnospiraceae*)	TTTN
LbCpf1 G532R/K595R	TYCV

As used herein, the expression “sgRNA” refers to a guide RNA which works in combination with a CRISPR nuclease to introduce a cut into DNA. The sgRNA comprises a sgRNA guide sequence and a “CRISPR nuclease recognition sequence”.

As used herein, the expression “sgRNA guide sequence” refers to the corresponding RNA sequence of the “sgRNA target sequence”. Therefore, it is the RNA sequence equivalent of the protospacer on the target polynucleotide gene sequence. It does not include the corresponding PAM sequence in the genomic DNA. It is the sequence that confers target specificity. The sgRNA guide sequence is linked to a CRISPR nuclease recognition sequence which binds to the nuclease (e.g., Cas9/Cpf1). The sgRNA guide sequence recognizes and binds to the targeted gene of interest. It hybridizes with (i.e., is complementary to) the opposite strand of a target gene sequence, which comprises the PAM (i.e., it hybridizes with the DNA strand opposite to the PAM). As noted above, the “PAM” is the nucleic acid sequence, that immediately follows (is contiguous to) the target sequence or target polynucleotide but is not in the sgRNA.

A “CRISPR nuclease recognition sequence” as used herein refers broadly to one or more RNA sequences (or RNA motifs) required for the binding and/or activity (including activation) of the CRISPR nuclease on the target gene. Some CRISPR nucleases require longer RNA sequences than other to function. Also, some CRISPR nucleases require multiple RNA sequences (motifs) to function while others only require a single short RNA sequence/motif. For example, Cas9 proteins require a tracrRNA sequence in addition to a crRNA sequence to function while Cpf1 only requires a crRNA sequence. Thus, unlike Cas9, which requires both crRNA sequence and a tracrRNA sequence (or a fusion or both crRNA and tracrRNA) to mediate interference, Cpf1 processes crRNA arrays independent of tracrRNA, and Cpf1-crRNA complexes alone cleave target DNA molecules, without the requirement for any additional RNA species (see Zetsche et al., PMID: 26422227).

The “CRISPR nuclease recognition sequence” included in the sgRNA described herein is thus selected based on the specific CRISPR nuclease used. It includes direct repeat sequences and any other RNA sequence known to be necessary for the selected CRISPR nuclease binding and/or activity. Various RNA sequences which can be fused to an RNA guide sequence to enable proper functioning of CRISPR nucleases (referred to herein as CRISPR nuclease recognition sequence) are well known in the art and can be used in accordance with the present disclosure. The “CRISPR nuclease recognition sequence” may thus include a crRNA sequence only (e.g., for AsCpf1 activity, such as the CRISPR nuclease recognition sequence UAAUUUCUAC UCUUGUAGAU (SEQ ID NO: 38)) or may include additional sequences (e.g., tracrRNA sequence necessary for Cas9 activity). Furthermore, in accordance with the present disclosure and as well known in the art, RNA motifs necessary for CRISPR nuclease binding and activity may be provided separately (e.g., (i) RNA guide sequence-crRNA CRISPR recognition sequence” (also known as crRNA) in one RNA molecule and (ii) a tracrRNA CRISPR recognition sequence on another, separate RNA molecule. Alternatively, all necessary RNA sequences (motifs) may be fused together in a single RNA guide. The CRISPR recognition sequence is preferably fused directly to the sgRNA guide sequence (in 3’ (e.g., Cas9) or 5’ (Cpf1) depending on the CRISPR nuclease used) but may include a spacer sequence separating two RNA motifs. In embodiments, the CRISPR nuclease recognition sequence is a Cas9 recognition sequence having at least 65 nucleotides. In embodiments, the CRISPR nuclease recognition sequence is a Cas9 CRISPR nuclease recognition sequence having at least 85 nucleotides. In embodiments, the CRISPR nuclease recognition sequence is a Cpf1 recognition sequence (5’ direct repeat) having about 19 nucleotides. In an embodiment, the CRISPR nuclease recognition sequence is a St1Cas9 recognition sequence. The sgRNA of the present disclosure may comprise any variant of the above noted sequences, provided that it allows for the proper functioning of the selected CRISPR nuclease (e.g., binding of the CRISPR nuclease protein to the gene of interest and/or target polynucleotide sequence(s)).

Together, the RNA guide sequence and CRISPR nuclease recognition sequence(s) provide both targeting specificity and scaffolding/binding ability for the CRISPR nuclease of the present disclosure. sgRNAs of the present disclosure do not exist in nature, i.e., is a non-naturally occurring nucleic acid(s).

A "target region", "target sequence" or "protospacer" in the context of sgRNAs and CRISPR system of the present disclosure are used herein interchangeably and refers to the region of the target gene, which is targeted by the CRISPR/nuclease-based system, without the PAM. It refers to the sequence corresponding to the nucleotides that

precede the PAM (i.e., in 5' or 3' of the PAM, depending of the CRISPR nuclease) in the genomic DNA. It is the sequence that is included into a sgRNA expression construct (e.g., vector/plasmid/AAV). The CRISPR/nuclease-based system may include at least one (i.e., one or more) sgRNAs, wherein each sgRNA target different DNA sequences on the target gene. The target DNA sequences may be overlapping. The target sequence or protospacer is followed or preceded by a PAM sequence at an (3' or 5' depending on the CRISPR nuclease used) end of the protospacer. Generally, the target sequence is immediately adjacent (i.e., is contiguous) to the PAM sequence (it is located on the 5' end of the PAM for SpCas9-like nuclease and at the 3' end for Cpf1-like nuclease).

In embodiments, the sgRNA of the present disclosure comprises a "sgRNA guide sequence" or has a "sgRNA target sequence" which corresponds to the target sequence on the gene of interest or target polynucleotide sequence that is followed or preceded by a PAM sequence (is adjacent to a PAM). The sgRNA may comprise a "G" at the 5' end of its polynucleotide sequence. The presence of a "G" in 5' is preferred when the sgRNA is expressed under the control of the U6 promoter (Taeyoung KooJungjoon Lee and Jin-Soo Kim Mol Cells. 2015 Jun 30; 38(6): 475–481). The CRISPR/nuclease system of the present disclosure may use sgRNAs of varying lengths. The sgRNA may comprise a sgRNA guide sequence of at least at least a 10, at least 12 nts, at least a 13 nts, at least a 14 nts, at least a 15 nts, at least a 16 nts, at least a 17 nts, at least a 18 nts, at least a 19 nts, at least a 20 nts, at least a 21 nts, at least a 22 nts, at least a 23 nts, at least a 24 nts, at least a 25 nts, at least a 30 nts, or at least a 35 nts of a target sequence of a gene of interest or target polynucleotide (such target sequence is followed or preceded by a PAM in the gene of interest or target polynucleotide but is not part of the sgRNA). The length of the sgRNA is selected based on the specific CRISPR nuclease used. In embodiments, the "sgRNA guide sequence" or "sgRNA target sequence" may be at least 17 nucleotides (17, 18, 19, 20, 21, 22, 23) long, preferably between 17 and 30 nts long, more preferably between 17-22 nucleotides long. In embodiments, the sgRNA guide sequence is between 10-40, 10-30, 12-30, 15-30, 18-30, or 10-22 nucleotides long. In embodiments, the PAM sequence is "NGG", where "N" can be any nucleotide. In embodiments, the PAM sequence is "TTTN", where "N" can be any nucleotide. sgRNAs may target any region of a target gene which is immediately adjacent (contiguous, adjoining, in 5' or 3') to a PAM (e.g., NGG/TTTN or CCN/NAAA for a PAM that would be located on the opposite strand) sequence. In embodiments, the sgRNA of the present disclosure has a target sequence that is located in an exon (the sgRNA guide sequence consists of the RNA sequence of the target (DNA) sequence which is located in an exon). In embodiments, the sgRNA of the present disclosure has a target sequence that is located in an intron (the sgRNA guide sequence consists of the RNA sequence of the target (DNA) sequence which is located in an intron). In embodiments, the sgRNA may target any region (sequence) which is followed (or preceded, depending on the CRISPR nuclease used) by a PAM in the gene or target polynucleotide of interest.

Although a perfect match between the sgRNA guide sequence and the DNA sequence on the targeted gene is preferred, a mismatch between a sgRNA guide sequence and target sequence on the gene sequence of interest is also permitted as along as it still allows hybridization of the sgRNA with the complementary strand of the sgRNA target polynucleotide sequence on the targeted gene. A seed sequence of between 8-12 consecutive nucleotides in the sgRNA, which perfectly matches a corresponding portion of the sgRNA target sequence is preferred for proper recognition of the target sequence. The remainder of the guide sequence may comprise one or more mismatches. In

general, sgRNA activity is inversely correlated with the number of mismatches. Preferably, the sgRNA of the present disclosure comprises 7 mismatches, 6 mismatches, 5 mismatches, 4 mismatches, 3 mismatches, more preferably 2 mismatches, or less, and even more preferably no mismatch, with the corresponding sgRNA target gene sequence (less the PAM). Preferably, the sgRNA nucleic acid sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% and 99% identical to the sgRNA target polynucleotide sequence in the gene of interest. Of course, the smaller the number of nucleotides in the sgRNA guide sequence the smaller the number of mismatches tolerated. The binding affinity is thought to depend on the sum of matching sgRNA-DNA combinations.

The number of sgRNAs administered to or expressed in a target cell in accordance with the methods of the present disclosure may be at least 1 sgRNA, at least 2 sgRNAs, at least 3 sgRNAs at least 4 sgRNAs, at least 5 sgRNAs, at least 6 sgRNAs, at least 7 sgRNAs, at least 8 sgRNAs, at least 9 sgRNAs, at least 10 sgRNAs, at least 11 sgRNAs, at least 12 sgRNAs, at least 13 sgRNAs, at least 14 sgRNAs, at least 15 sgRNAs, at least 16 sgRNAs, at least 17 sgRNAs, or at least 18 sgRNAs. The number of sgRNAs administered to or expressed in a cell may be between at least 1 sgRNA and 15 sgRNAs, 1 sgRNA and least 10 sgRNAs, 1 sgRNA and 8 sgRNAs, 1 sgRNA and 6 sgRNAs, 1 sgRNA and 4 sgRNAs, 1 sgRNA and sgRNAs, 2 sgRNA and 5 sgRNAs, or 2 sgRNAs and 3 sgRNAs.

CRISPR nucleases

Recombinant dCas9-FoKI dimeric nucleases (RFNs) have been designed that can recognize extended sequences and edit endogenous genes with high efficiency in human cells. These nucleases comprise a dimerization-dependent wild type FokI nuclease domain fused to a catalytically inactive Cas9 (dCas9) protein. Dimers of the fusion proteins mediate sequence specific DNA cleavage when bound to target sites composed of two half-sites (each bound to a dCas9 (i.e., a Cas9 nuclease devoid of nuclease activity) monomer domain) with a spacer sequence between them. The dCas9-FoKI dimeric nucleases require dimerization for efficient genome editing activity and thus, use two sgRNAs for introducing a cut into DNA.

The recombinant CRISPR nuclease that may be used in accordance with the present disclosure is i) derived from a naturally occurring Cas; and ii) has a nuclease (or nickase) activity to introduce a DSB (or two SSBs in the case of a nickase) in cellular DNA when in the presence of appropriate sgRNA(s). Thus, as used herein, the term “CRISPR nuclease” refers to a recombinant protein which is derived from a naturally occurring Cas nuclease which has nuclease or nickase activity and which functions with the sgRNAs of the present disclosure to introduce DSBs (or one or two SSBs) in the targets of interest. In an embodiment, the CRISPR nuclease is St1Cas9. In further embodiments, the CRISPR nuclease is SpCas9 or Cpf1. In another embodiment, the CRISPR nuclease is a Cas9 protein having a nickase activity. As used herein, the term “Cas9 nickase” refers to a recombinant protein which is derived from a naturally occurring Cas9 and which has one of the two nuclease domains inactivated such that it introduces single stranded breaks (SSB) into the DNA. It can be either the RuvC or HNH domain. In a further embodiment, the Cas protein is a dCas9 protein fused with a dimerization-dependant FokI nuclease domain.

Exemplary CRISPR nucleases that may be used in accordance with the present disclosure are provided in Table 1 above.

CRISPR nucleases such as Cas9/nucleases cut 3-4bp upstream of the PAM sequence. CRISPR nucleases such as Cpf1 on the other hand, generate a 5' overhang. The cut occurs 19 bp after the PAM on the targeted (+) strand and 23 bp on the opposite strand (Zetsche et al., 2015, PMID 26422227). There can be some off-target DSBs using wildtype Cas9. The degree of off-target effects depends on a number of factors, including: how closely homologous the off-target sites are compared to the on-target site, the specific site sequence, and the concentration of nuclease and guide RNA (sgRNA). These considerations only matter if the PAM sequence is immediately adjacent to the nearly homologous target sites. The mere presence of additional PAM sequences should not be sufficient to generate off target DSBs; there needs to be extensive homology of the protospacer followed or preceded by PAM.

Optimization of codon degeneracy

Because CRISPR nuclease proteins are (or are derived from) proteins normally expressed in bacteria, it may be advantageous to modify their nucleic acid sequences for optimal expression in eukaryotic cells (e.g., mammalian cells) when designing and preparing CRISPR nuclease recombinant proteins. Similarly, donor or patch nucleic acids of the present disclosure used to introduce specific modifications in the target polynucleotide may use codon degeneracy (e.g., to introduce new restriction sites for enabling easier detection of the targeted modification).

Accordingly, the following codon chart (Table 2) may be used, in a site-directed mutagenic scheme, to produce nucleic acids encoding the same or slightly different amino acid sequences of a given nucleic acid:

Table 2: Codons encoding the same amino acid

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU

Amino Acids			Codons					
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

MODE(S) FOR CARRYING OUT THE INVENTION

The present disclosure is illustrated in further details by the following non-limiting examples.

Example 1: Materials and methods

Cell culture and transfection

K562 were obtained from the ATCC (CCL-243) and maintained at 37 °C under 5% CO_2 in RPMI medium supplemented with 10% FBS, penicillin-streptomycin and GlutaMAX™. Neuro-2a were obtained from the ATCC and maintained at 37 °C under 5% CO2 in DMEM medium supplemented with 10% FBS, penicillin-streptomycin and GlutaMAX™. All cell lines are tested for absence of mycoplasma contamination. Cells ($2x10^5$ per transfection) were transfected using the Amaxa 4D-Nucleofector (Lonza) per manufacturer's recommendations. K562 cell lines expressing SaCas9 and St1Cas9 from the *AAVS1* safe harbor locus were generated as described[35, 36]. Briefly, simultaneous selection and cloning was performed for 10 days in methylcellulose-based semi-solid RPMI medium supplemented with 0.5 µg/ml puromycin starting 3 days post-transfection. Clones were picked and expanded in 96 wells for 3 days and transferred to 12-well plates for another 3 days before cells were harvested for western blot.

Genome editing vectors

Vectors for *in vitro* and *in vivo* genome editing with the CRISPR1-StCas9 LMD-9 system generated in this study are available from Addgene (**Fig. 11**). The CRISPOR[39] web tool was used to design guide (spacer) sequences against mouse and human targets (**Tables 3-6)**. DNA sequence for the spacers were modified at position 1 to encode a "G" due to the transcription initiation requirement of the human U6 promoter when required. Alternatively, the spacer length was increased to capture a naturally occurring "G". The mammalian expression vector for *S. thermophilus* CRISPR1 (St1Cas9 LMD-9) fused to SV40 NLS sequences at the N- and C-terminus (MSP1594_2x_NLS; Addgene plasmid #110625) was constructed from MSP1594[34] (Addgene plasmid #65775). The U6-driven sgRNA expression plasmids for *S. thermophilus* CRISPR1 (St1Cas9 LMD-9) (v1) (St1Cas9_LMD-9_sgRNA_pUC19; Addgene plasmid #110627) and SaCas9[7] were synthesized as gBlocks™ gene fragments (Integrated DNA Technologies) and cloned into pUC19. BPK2301[34] (v0) (Addgene plasmid # 65778) was used to compare St1Cas9 sgRNA architectures. The single vector mammalian expression system containing a CAG promoter-driven St1Cas9 LMD-9 and its U6-driven sgRNA (U6_sgRNA_CAG_hSt1Cas9_LMD9; Addgene plasmid #110626) was built from the above-described plasmids. The single vector rAAV-St1Cas9 LMD-9 systems containing liver-specific promoters (**Table 8**) were

assembled from the above-described components into a derivative of pX602[7] (Addgene plasmid #61593) containing a deletion within the backbone to eliminate BsmBI restriction sites. The LP1b promoter was engineered by combining elements from previously described AAV expression cassettes[53, 54]. The most active version of this vector (v3) has the structure pAAV_LP1B_St1Cas9_LMD-9_SpA_U6_sgRNA (Addgene plasmid # 110624). To establish clonal K562 cell lines constitutively expressing C-terminally tagged SaCas9 and St1Cas9 under the control of an h*PGK1* promoter, the Cas9 ORFs from pX602 and MSP1594_2x_NLS were subcloned into AAVS1_Puro_PGK1_3xFLAG_Twin_Strep[36] (Addgene plasmid # 68375).

Surveyor nuclease and TIDE assays

Genomic DNA from 2.5E5 cells was extracted with 250 ml of QuickExtract™ DNA extraction solution (Epicentre) per the manufacturer's recommendations. The various loci were amplified by PCR using the primers described in **Table 9**. Assays were performed with the Surveyor™ mutation detection kit (Transgenomics) as described[35, 37]. Samples were separated on 10% PAGE gels in TBE buffer. Gels were imaged using a ChemiDoc™ MP (Bio-Rad) system and quantifications were performed using Image Lab™ software (Bio-Rad). TIDE analysis was performed using a significance cut-off value for decomposition of $p<0.001$[38].

Recombinant adeno-associated virus production

Production of recombinant adeno-associated viral vectors was performed by the triple plasmid transfection method essentially as described[81]. Briefly, HEK293T17 cells were transfected using polyethylenimine (PEI, Polysciences) with helper plasmid pxx-680, the rep/cap hybrid plasmid pAAV2/8 and the rAAV vector plasmid. Twenty-four hours post-transfection, media was replaced with growth media without FBS, and cells were harvested 24 hours later. AAV particles were extracted from cell extracts by freeze/thaw cycles and purified on a discontinuous iodixanol gradient. Virus were resuspended in PBS 320mM NaCl + 10% sorbitol + 0.002% pluronic acid, aliquoted and stored at −80°C. AAV were titrated by qPCR (Roche) using SYBR™ green and ITR primers as described[82]. Physical titer and purity was confirmed by separating similar volumes of AAV on a 10% SDS-PAGE stain free gel (Biorad) in Tris-Glycine-SDS buffer. ITR integrity was assessed following a BssHII digestion of the AAV plasmid. The vector core facility at the CERVO brain research center (Université Laval) produced the rAAV8s.

Animal experiments

Fah$^{-/-}$ mice[83] on a C57BL/6 genetic background were group-housed and fed a standard chow diet (Harlan #2018SX) with free access to food and water. *Fah*$^{-/-}$ mice drinking water was supplemented with 7.5 mg (2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione) (NTBC)/L and pH was adjusted to 7.0. Mice were exposed to a 12:12-h dark-light cycle and kept under an ambient temperature of 23 ± 1 °C. Animals were cared for and handled according to the *Canadian Guide for the Care and Use of Laboratory Animals*. The Université Laval Animal Care and Use Committee approved the procedures.

Two days old neonatal mice were injected intravenously in the retro-orbital sinus[84] with different doses of rAAV8 or saline in a total volume of 20 µL. Mice were weaned at 21 days of age and NTBC was removed 7 days later. Body weight and glycemia were monitored daily following NTBC removal. Mice were not fasted for measurement of

glycemia, data collection occurred between 9-10 am. Animals were killed by cardiac puncture under anesthesia at predetermined time points or when weight loss reached 20% of body weight. Livers were snap frozen for downstream applications.

Urine collection and succinylacetone quantification

Urine from groups of 3-4 mice was collected overnight in metabolic cages (Tecniplast) 15 days after NTBC removal. Urine was centrifuged at 2000 rpm for 5 minutes, aliquoted and frozen at -80°C. Succinylacetone was quantified in urine samples by a sensitive method using gas chromatography–mass spectrometry (GC-MS) as previously described[85]. The biochemical genetics laboratory at the Centre Hospitalier universitaire de Sherbrooke performed the analyses.

Example 2: Identification of an sgRNA architecture directing robust DNA cleavage by St1Cas9 in human cells

S. thermophilus encodes up to two type II-A systems (CRISPR1 and CRISPR3). While characterizing the interplay between St1Cas9 and diverse Acr families isolated from phages infecting *S. thermophilus*[32], we were surprised by the substantial levels of editing achieved in human cells. This observation contrasts with early reports indicating that this ortholog was mildly active[7, 33].

In the studies described herein, we made various modifications which we found were capable of increasing activity. First, we added an N-terminal nuclear localization signal (NLS) to a human codon-optimized expression construct[34] and established a K562 cell line stably expressing St1Cas9 (LMD-9) from the *AAVS1* safe harbor locus[35, 36] (**Figs. 1a and 5**). St1Cas9 (1121 aa) shares 17% and 37% identity with SpCas9 and SaCas9, respectively. Second, we adapted an sgRNA sequence used to monitor St1Cas9 activity in the heterologous host *Escherichia coli*[31]. We substituted a wobble base pair present in the lower stem of the repeat:anti-repeat region for a canonical Watson-Crick base pair in order to interrupt the RNA polymerase III termination signal (Fig. 1b). Then, we compared this sgRNA architecture (v1) to its counterpart containing a wild-type full length crRNA:tracrRNA duplex connected via a tetraloop (v0) by targeting *EMX1*, *FANCF*, and *RUNX1*[34] (**Figs. 1c and 5**). St1Cas9-expressing cells were transfected with increasing amounts of each construct and the Surveyor nuclease assay was used to determine the frequency of indels characteristic of imprecise DSB repair by NHEJ[35, 37] (**Fig. 1d**). The spectrum and frequency of targeted mutations was also analyzed using the complementary TIDE (Tracking of Indels by DEcomposition) method[38] (Fig. 1d and Table 3). Irrespective of the quantification method, the potency of sgRNA v1 was markedly superior. The increased activity was also observed when co-expressing St1Cas9 and its sgRNA transiently, a setting more typical of a genome editing experiment (**Fig. 5**). This analysis revealed that high gene disruption rates could be obtained under standard conditions using St1Cas9 in human cells.

Example 3: Robust editing of target genes involved in liver metabolism by St1Cas9 in mouse cells

We used CRISPOR[39] to design sgRNAs against *Pck1*, *Pcsk9,* and *Hpd*, three genes affecting liver function when disrupted. When possible, we selected guides targeting essential protein domains and predicted to have few

potential off-targets. Transient transfection of single vector constructs expressing both St1Cas9 and its sgRNA revealed strong cleavage activity (18% to > 50% indels) at 14 out of 15 target sites highlighting the robustness of the system despite not relying on sgRNA design rules[33] (**Fig. 2a-c**). Of note, this screen identified highly active sgRNAs targeting in the vicinity of mutations found in human *HPD*[40, 41]. Deficiency of 4-hydroxyphenyl-pyruvate dioxygenase (*HPD*), the second enzyme in the tyrosine catabolic pathway, causes Tyrosinemia type III (Orphanet ORPHA:69723) (**Fig. 3a**). Only three missense mutations are known to cause this rare disease (Prevalence <1/1,000,000) and we could target two of them with high efficacy (OMIM 276710) (**Figs. 3c** and **6**). Targeting the third mutation was not attempted due to the low specificity score of the guide. Taken together, these data suggest that St1Cas9 might enable *in vivo* genome editing if it could be packaged into a single rAAV particle alongside its sgRNA and the regulatory elements needed to drive its expression.

Example 4: Potent *in vivo* genome editing using an all-in-one rAAV vector in newborn mice

To deliver holo-St1Cas9 (St1Cas9 + sgRNA) to the liver, we generated a hepatotropic rAAV serotype 8[11, 16-18] vector targeting *Hpd* exon 13 (aka AAV8-St1Cas9 *Hpd* G5) by mirroring the original SaCas9 vector architecture[7] (**Fig. 3c**). To test the cleavage activity of St1Cas9 *in vivo*, we injected mice at day 2 of life into the retro-orbital sinus with increasing amounts of vector and isolated total liver DNA at day 28 post injection (**Fig. 3b**). The titration showed that the degree of editing was substantial and dependent on the dose of AAV8-St1Cas9 (**Fig. 3d**).

To test if AAV8-St1Cas9 can lead to phenotypic correction *in vivo*, we used a mouse model of hereditary tyrosinemia type I (HT-I) (OMIM 276700) (Orphanet ORPHA:882), an autosomal recessive disease caused by a deficiency of fumarylacetoacetate hydrolase (FAH), the last enzyme of the tyrosine catabolic pathway (**Fig. 3a**). Of particular relevance to us, the incidence of HT-I reaches 1/1846 in a region of the province of Québec (Canada) while it is around 1/100,000 births worldwide[49]. *Fah*$^{-/-}$ mutant mice die as neonates with severe hepatic dysfunction and kidney damage due to the accumulation of toxic metabolites unless treated with nitisone (NTBC), a drug that inhibits *Hpd* upstream in the pathway (**Fig. 3a**)[50]. Similarly, genetic ablation of *Hpd* prevents liver damage and lethality [51, 52]. *Fah*$^{-/-}$ mutant pups maintained on NTBC were injected at day 2 of life with AAV8-St1Cas9 *Hpd* G5 and then the drug was withdrawn shortly after weaning (**Fig. 3b**). Systemic delivery via a single neonatal injection rescued lethality in all mice while saline-treated animals had to be killed after ~3 weeks as they lost weight (**Fig. 3e,f**). Likewise, glycemia was normalized in the treatment groups (**Fig. 3g**). Notably, the excretion of succinylacetone, a toxic metabolite and a diagnostic marker for HT-I, was inversely correlated with the dose of rAAV demonstrating metabolic correction (**Fig. 3h**). These observations were recapitulated when targeting *Hpd* exon 8 at a site corresponding to a mutation also found in human patients (**Fig. 6**). Therefore, rAAV-mediated delivery of St1Cas9 *in vivo* can correct a phenotype in neonatal mice by rewiring a metabolic pathway.

Lastly, we evaluated two additional vector architectures in order to minimize the size of rAAV and test the impact of the promoter on overall activity (**Fig. 4**). An rAAV vector (v3) containing an engineered liver-specific promoter (LP1b) combining the human apolipoprotein E/C-I gene locus control region (*ApoE*-HCR) and a modified human α1 antitrypsin promoter (h*AAT*) coupled to an SV40 intron and a synthetic polyadenylation element greatly improved

efficacy as compared to the TBG promoter (**Fig. 4a,b**). These modifications also led to the creation of a vector of ~4.7 kb in size which was optimal for efficient packaging. Collectively, these data indicate that St1Cas9 is an efficient tool for *in vivo* genome editing.

It is shown herein that St1Cas9 can be harnessed for robust and efficient genome editing *in vitro* and *in vivo*. While there is considerable interest in exploiting the diversity of Cas enzymes, but their implementation as genome editing tools is not a straightforward process[7-10]. Some enzymes simply fail to work and some choose their substrates promiscuously, necessitating thorough biochemical characterization[58-64]. Moreover, sgRNAs for St1Cas9 and SaCas9 are not functionally interchangeable, which is likely due to their unique PAM specificity (**Fig. 7**).

Cas9 orthologs used for rAAV-mediated *in vivo* genome editing require a more complex PAM than the relatively simple NGG of SpCas9. This restricts the range of accessible targets but may increase specificity by reducing the occurrence of off-target mutagenesis. The consensus PAM for St1Cas9 (LMD-9 and DGCC7710 strains that differ by only 2 aa) has been defined as $N^1N^2A^3G^4A^5A^6(W^7)$, however sequences closely related to the consensus can be functional in test tubes and in bacterial cells[29, 34, 73-76]. While recognition of an A-rich PAM may ease targeting A/T-rich regions of genomes, we found that St1Cas9 can be targeted to both NNAGAA and NNGGAA PAMs in mammalian cells (**Fig. 8**). Of note, the presence of an A at position 7 of the PAM correlates with high activity (**Fig. 8**). While the length of the nonconserved linker (N^1N^2) has also been shown to be flexible and an extension from 2 to 3 bases has been shown to be tolerated[31, 77], we failed to reproduce this observation in human cells suggesting a higher stringency of the system in this context (**Fig. 8**).

In embodiments, different St1Cas9 PAM sequences may be used, for example, inferred consensus PAM sequences for St1Cas9 from strains CNRZ1066 and LMG13811 are NNACAA(W) and NNGCAA(A), respectively[24, 26]. Notably, LMG13811 CRISPR1 system transplanted in *E. coli* or reconstituted from purified components can target DNA using the NNGCAAA PAM[77]. At the protein level, the sequence of these three St1Cas9 variants diverges mostly within the C-terminal PAM-interacting (PI) domain (**Figs. 9-10**).

Example 5: Engineering St1Cas9 nucleases with altered protospacer adjacent motif (PAM) specificities.

One constraint for the use of St1Cas9 is its requirement for a longer PAM of the form $N_1N_2A_3G_4A_5A_6W_7$ (where W is A or T) that can restrict targeting. This consensus was initially obtained by examining the sequences flanking CRISPR-Cas9 target sites within bacteriophage genomes. However, sequences closely related to the consensus (NNAGAAW and NNGGAAW) can be functional in test tubes or when transplanted in *E. coli*. These differences are believed to emerge from the different stringency imparted by the heterologous systems. Nevertheless, these deviations from the consensus suggest that there is some flexibility in PAM recognition. Thus, it is crucial to define functional PAMs for each Cas9 in their proper context; in our case, human and mouse cells.

We first codon-optimized St1Cas9 for expression in human cells and appended N- and C-terminal nuclear localization signals (NLS). We show that both NNAGAAW_7 and NNGGAAW_7 PAM sequences could direct DNA cleavage with equivalent efficacy in cells (**Fig. 18**). We also observed that substitutions at position 7 are well tolerated. (**Fig. 18**). Thus, it appears that a functional PAM for St1Cas9 requires a core of four specific base pairs (NNAGAA or NNGGAA). By itself, removing the requirement for a W at position 7, increases the targeting range twofold. Since

NNGGAA PAMs behave similarly, this results in an additional twofold expansion. By comparison, SaCas9 requires an NNGRRT (where R is A or G) PAM for cleavage.

We show herein the identification of St1Cas9 enzymes with distinct PAM specificities. The St1Cas9 protein sequence used in nearly all studies so far was derived from the LMD-9 or DGCC7710 strains that differ by only two conservative substitutions. We studied the LMD-9 St1Cas9 as well as St1Cas9 from strains LMG18311, CNRZ1066 and TH1477. As noted above, at the protein level, the sequence of these three St1Cas9 variants diverges mostly within the C-terminal wedge (WED) and PAM-interacting (PI) domains (**Figs. 10 and 35b**). Using the structure of SaCas9[28] as a guide, we tested whether swapping the C-terminus of St1Cas9 LMD-9 with the ones from the LMG 18311, CNRZ 1066 and TH1477 could reprogram PAM specificity, and we thus engineered hybrid proteins containing the N-terminal domain of St1Cas9 LMD-9 (REC lobe, HNH and RuvC nuclease domains, and phosphate lock loop ; aa 1-826) and C-terminal domains of St1Cas9 LMG 18311, CNRZ 1066 and TH1477 (WED and PI domains ; aa 827-1121) (**Fig. 19**). While St1Cas9 LMD-9 could only target NNAGAA and NNGGAA PAMs, the hybrid constructs targeted with high efficacy NNACAA and NNGCAA PAMs, respectively (**Fig. 19**). We observed limited cross reactivity indicating that true reprogramming, as opposed to relaxed specificity was achieved. These data highlight the modularity inherent to Cas9 enzymes, and this strategy may be used to further expand the targeting range of St1Cas9.

Example 6: Engineering St1Cas9 variants with expanded targeting range

In an effort to identify additional St1Cas9 proteins with novel PAM requirements, we used a recently published bioinformatics pipeline called “Search for PAMs by ALignment Of Targets” (SPAMALOT)[86]. This process identified an additional St1Cas9 represented by strain TH1477 that potentially targets NNGAAA PAMs (**Fig. 20a**). We generated chimeric fusion proteins with the N-terminus of LMD-9 and the C-terminal domain of TH1477 as we have done for CNRZ1066 and LMG1831. This approach yielded an active St1Cas9 derived from the TH1477 strain that can target NNGAAA PAMs (**Fig. 20b**).

Example 7: Converting St1Cas9 to a base editor.

DNA base editors comprise fusions between a catalytically impaired Cas nuclease and a base modification enzyme that operates on single-stranded DNA (ssDNA)[80]. Cytosine base editors (CBEs) convert a C•G base pair into a T•A using the APOBEC1 cytidine deaminase. Fusion of APOBEC1 to the *Streptococcus pyogenes* D10A mutant (nickase) and two copies of the uracil DNA glycosylase inhibitor (UGI), resulted in the creation of BE4max enzyme. The *Staphylococcus aureus* Cas9 has also been converted into a base editor to create SaBE4. We have created St1BE4max by exchanging SpCas9 D10A for St1Cas9 D9A (LMD-9) into the original BE4max construct. This created a potent CBE with novel targeting specificity due to the unique PAM of St1Cas9 (**Fig. 21**). Our data indicate that St1BE4max has a similar activity window to SaBE4. Since the activity window (aka editing window) of base editors is narrow there is a distinct advantage of creating base editors targetable to a broad range of PAM sequences. This is particularly important considering the recent engineering of deaminase domains with even more narrower editing

windows, such as APOBEC3A (eA3A), which preferentially deaminates cytidines in specific motifs according to a TCR>TCY>VCN hierarchy[80].

We then proceeded to demonstrate that the St1Cas9 strain variants that display unique PAM preferences are also functional as CBEs. Specifically, LMD-9/LMG18311 hybrid- and LMD-9/CNRZ1066 hybrid-based St1BE4max are potent base editors at NNGCAA and NNACAA PAMs, respectively (**Fig. 22**). These data further demonstrate the use of St1Cas9 as a genome editing platform and the value of creating St1Cas9 fusions based on variants.

Table 3: St1Cas9 guide (spacer) sequences targeting NNAGAA PAMs (Examples 1-4)

Gene	ID	bp	5'	Target		PAM	3'
				Sequence	SEQ ID NO:		
EMX1	G1	22	ACCTGG	GCCAGGGAGGGAGGGGCACAGA	70	TGAGAA	ACTCAG
	G2	19	AGAACC	GGAGGACAAAGTACAAACG	71	GCAGAA	GCTGGA
	G3	24	GTTCCA	GAACCGGAGGACAAAGTACAAACG	72	GCAGAA	GCTGGA
	G4	20	CTGGAG	GAGGAAGGGCCTGAGTCCGA	73	GCAGAA	GAAGAA
	G5	20	GAGGAG	GAAGGGCCTGAGTCCGAGCA	74	GAAGAA	GAAGGG
	G6	20	GAGGAA	GGGCCTGAGTCCGAGCAGAA	75	GAAGAA	GGGCTC
	G7	23	GGGCCC	GCCCAGGCAGGCAGGCTCTCCGA	76	GGAGAA	GGCCAA
FANCF	G1	20	AAGCTC	GGAAAAGCGATCCAGGTGCT	77	GCAGAA	GGGATT
	G2	20	GCTGAC	GTAGGTAGTGCTTGAGACCG	78	CCAGAA	GCTCGG
RUNX1	G1	21	ATTACT	GTACTAATCAGATGGAAGCTCT	79	TCAGAA	ATGTTT
	G2	22	GTAAAA	GAAATCATTGAGTCCCCCGCCT	80	TCAGAA	GTGGGT
	G3	20	GTCCCT	GAGGTATCCAGCAGAGGGGA	81	GAAGAA	AGAGAG
	G4	23	TGGGGA	GTCCCAGAGGTATCCAGCAGAGG	82	GGAGAA	GAAAGA
ATP1A1	G1	20	TCTGTA	GCAGCTTGGATGCTATAAGC	83	CAAGAA	ACAAAG
Hpd	G1	20	CTACTT	GGTACCCCACGCAGAAAGCT	84	CGAGAA	CGGGGC
	G2	20	TTGTAT	GTTGGGGCCTCGAATCCAGG	85	TAAGAA	ACGGCC
	G3	20	TATGGA	GATACCACACACACCCTGGT	86	GGAGAA	GATCAA
	G4	21	CAGTTT	GTAGTAAGAAGATGGGGCGGC	87	CAAGAA	CTCCGT
	G5	21	GGAGCT	GCATATCCTAGTCGACTATGA	88	CGAGAA	AGGCTA
Pcsk9	G1	19	CCAACA	GGTCACTGCTCATCTTCAC	89	CAAGAA	GCCAGG
	G2	19	CCAACA	GGGTCACTGCTCATCTTCAC	90	CAAGAA	GCCAGG
	G3	20	CCCAAC	GAGGTCACTGCTCATCTTCAC	91	CAAGAA	GCCAGG
	G4	19	AATCAC	GCACGACGCCTCCCGCTCCT	92	GGAGAA	GCTGGA
	G5	20	CAATCA	GCCACGACGCCTCCCGCTCCT	93	GGAGAA	GCTGGA
	G6	20	GGCCTG	GAGACCCATGTCCACTGCCA	94	CCAGAA	GGACCA
Pck1	G1	20	GGATAT	GGTGGGAACTCACTACTCGG	95	GAAGAA	ATGCTT
	G2	20	ATCCTG	GGCATAACTAACCCCGAAGG	96	CAAGAA	GAAATA
	G3	20	ATAATG	GGGCACTGGCTGGCAGGGGT	97	GCAGAA	TCTCGA

	G4	20	GCCAGG	GTATTTGCCGAAGTTGTAGCC	98	GAAGAA	GGGTCG

"G" in position 1 of the guide indicates a mismatch to the genome and it is not counted in the size (bp) of the guide.

Table 4: St1Cas9 guide target sequences targeting NNGGAA PAMs (Examples 1-4)

Gene	ID	bp	5'	Target		PAM	3'
				Sequence	SEQ ID NO:		
EMX1	**G8**	22	GACAAA	GTACAAACGGCAGAAGCTGGAG	99	GAGGAA	GGGCCT
FANCF	**G3**	20	GCGGAA	GTAGGGCCTTCGCGCACCTC	100	ATGGAA	TCCCTT
	G4	19	GGTAGT	GCTTGAGACCGCCAGAAGC	101	TCGGAA	AAGCGA
	G5	21	TAGGTA	GTGCTTGAGACCGCCAGAAGC	102	TCGGAA	AAGCGA
	G6	20	ACCGAG	GGCCTGGAAGTTCGCTAATC	103	CCGGAA	CTGGAC
VEGFA	**G1**	20	GGTGGG	GAGAGGGACACACAGATCTA	104	TTGGAA	TCCTGG
	G2	20	GGGCCT	GAGAGCCGTTCCCTCTTTGC	105	TAGGAA	TATTGA
	G3	20	CCCGCG	GGGCATTGGCGAGGAGGGAG	106	CAGGAA	AGTGAG
	G4	19	CAGCCT	GAAAATTACCCATCCGCCC	107	CCGGAA	ACTCTG
	G5	23	TTCACA	GCCTGAAAATTACCCATCCGCCC	108	CCGGAA	ACTCTG

Table 5: St1Cas9 guide target sequences targeting PAMs with a NNN linker (Examples 1-4)

Gene	ID	bp	5'	Target		PAM	3'
				Sequence	SEQ ID NO:		
EMX1	**G5**	20	GGAGGA	GGAAGGGCCTGAGTCCGAGC	109	AGAAGAA	GAAGGG
FANCF	**G2**	19	GCTGAC	GTAGGTAGTGCTTGAGACC	110	GCCAGAA	GCTCGG
	G5	20	TAGGTA	GTGCTTGAGACCGCCAGAAG	111	CTCGGAA	AAGCGA
ATP1A1	**G1**	19	TCTGTA	GCAGCTTGGATGCTATAAG	112	CCAAGAA	ACAAAG
RUNX1	**G3**	19	GTCCCT	GAGGTATCCAGCAGAGGGG	113	AGAAGAA	AGAGAG

Table 6: SaCas9 guide (spacer) sequences targeting NNGRRT PAMs (Examples 1-4)

Gene	ID	bp	5'	Target		PAM	3'
				Sequence	SEQ ID NO:		
EMX1	**G1**	21	GGGTGG	GCAACCACAAACCCACGAGGG	114	CAGAGT	GCTGCT

FANCF	**G1**	21	GCGGAA	GTAGGGCCTTCGCGCACCTCA	115	TGGAAT	CCCTTC
RUNX1	**G1**	23	CAGCAT	GTACTCACCTCTCATGAAGCACT	116	GTGGGT	ACGAAG
Hgd	**G1**	20	CATCCT	GGAGGTCTATGGTGTCCACT	117	TTGAGT	TACCTG
Hpd	**G1**	20	AGGTGA	GAGTTTGCTGTGCTGCAGACG	118	GTGAGT	GAACAC

"G" in position 1 of the guide indicates a mismatch to the genome and it is not counted in the size (bp) of the guide.

Table 7: SaCas9 and St1Cas9 sgRNAs

sgRNA	Sequence
St1Cas9_v0	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACGAAACTTACACAGTTACTTAAATCTTGCAGAAGCTACA AAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT (SEQ ID NO: 119) GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAACGAAACUUACACAGUUACUUAAAUCUUGCAGAAG CUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU (RNA; SEQ ID NO: 120)
St1Cas9_v1	GTCTTTGTACTCTGGTACCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGG CAGGGTGTTTT (SEQ ID NO: 121) GUCUUUGUACUCUGGUACCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUU AUGGCAGGGUGUUUU (RNA; SEQ ID NO: 122)
SaCas9	GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCG AGAT (SEQ ID NO: 123 GUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUU GGCGAGAU (RNA; SEQ ID NO: 124)

Table 8: Sequences for TBG and LP1B promoters

Promoter	Sequence

TBG	GGGCTGGAAGCTACCTTTGACATCATTTCCTCTGCGAATGCATGTATAATTTCTACAGAACCTATTAGAAAGG ATCACCCAGCCTCTGCTTTTGTACAACTTTCCCTTAAAAAACTGCCAATTCCACTGCTGTTTGGCCCAATAGT GAGAACTTTTTCCTGCTGCCTCTTGGTGCTTTTGCCTATGGCCCCTATTCTGCCTGCTGAAGACACTCTTGC CAGCATGGACTTAAACCCCTCCAGCTCTGACAATCCTCTTTCTCTTTTGTTTTACATGAAGGGTCTGGCAGCC AAAGCAATCACTCAAAGTTCAAACCTTATCATTTTTTGCTTTGTTCCTCTTGGCCTTGGTTTTGTACATCAGCT TTGAAAATACCATCCCAGGGTTAATGCTGGGGTTAATTTATAACTAAGAGTGCTCTAGTTTTGCAATACAGGA CATGCTATAAAAATGGAAAGAT (SEQ ID NO: 125)
LP1b	CCCTAAAATGGGCAAACATTGCAAGCAGCAAACAGCAAACACACAGCCCTCCCTGCCTGCTGACCTTGGAG CTGGGGCAGAGGTCAGAGACCTCTCTGGGCCCATGCCACCTCCAACATCCACTCGACCCCTTGGAATTTCG GTGGAGAGGAGCAGAGGTTGTCCTGGCGTGGTTTAGGTAGTGTGAGAGGGTGGACACAGGACGCTGTGGT TTCTGAGCCAGGGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCCTGTTTGCTCCTCCGATAACTGGGG TGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGTTGCCCCTCTGGATCCACTGCTTAAATACGGACGAGG ACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTGGGACAGTGAATCCGGACTCTAAGGTAAA TATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTCTCTCTTTTAGATTCCAACCTTTGG AACTGA (SEQ ID NO: 126)

Table 9: PCR primers and amplicon sizes for Surveyor and TIDE assays.

Target	Primer	SEQ ID NO:	Size (bp)
EMX1 Forward	CCATCCCCTTCTGTGAATGT	127	639
EMX1 Reverse	GGAGATTGGAGACACGGAGA	128	
FANCF Forward	AGGAACACGGATAAAGACGCTGG	129	492
FANCF Reverse	AGTTGCTGCACCAGGTGGTAACG	130	
RUNX1 Forward	CCAGCACAACTTACTCGCACTTGAC	131	601
RUNX1 Reverse	CATCACCAACCCACAGCCAAGG	132	
VEGFA Forward	GAGAAGGCCAGGGGTCACTCCAG	133	278
VEGFA Reverse	AGCCCGCCGCAATGAAGG	134	

Hpd exon 7 Forward	GCAGGCGCAGTGCCCAAGACAC	135	498
Hpd exon 7 Reverse	CAGCACATGCCCAGGTCACATGG	136	
Hpd exon 8 Forward	GCCATGAGGACAGAAAGAGCATC	137	409
Hpd exon 8 Reverse	GATATTCCAGTCTCCCAGAGAAG	138	
Hpd exon 12 Forward	CTCGCATACTTGAAGGCTGTGCC	139	429
Hpd exon 12 Reverse	GATAGGGACTCTGCTACCTCCTG	140	
Hpd exon 13 Forward	GGCTTTGGTGGTGCAGTAGCCTT	141	402
Hpd exon 13 Reverse	GACCTCACACCATTGGGCTCCAG	142	
Pck1 exon 5 Forward	GATGTGCACCACACAAGCTCACTGT	143	567
Pck1 exon 5 Reverse	GAGTTCGGTGCTGTTGTCTAAGA	144	
Pck1 exon 6 Forward	GCTAGTTTGGAAGACAGTCCTAG	145	494
Pck1 exon 6 Reverse	GTCCCTCTCTATCCAGATGATCC	146	
Pck1 exon 8 Forward	GCAACTTAAGGGCTATCAACCCA	147	556
Pck1 exon 8 Reverse	GTCTGGATATAGGAGGGAGATCT	148	
Pck1 exon 10 Forward	GAACACAAGGGTGAGTCACAGTC	149	565
Pck1 exon 10 Reverse	CATCTGGCTGATTCTCTGTTTCA	150	
Pcsk9 exon 2 Forward	GATACGCATGCTACACTGAGATG	151	534
Pcsk9 exon 2 Reverse	CACAGGCAGTAGACAAACCAG	152	
Pcsk9 exon 9 Forward	GCACCCATGAGACAGGTGAGCAG	153	559

Pcsk9 exon 9 Reverse	GCAGAGACAATGGGTGGCTAATA	154	
Pcsk9 exon 10 Forward	GCTGGAAGCTTTATGATGGAGAT	155	599
Pcsk9 exon 10 Reverse	GACACACCTCAGAGCCTTCCCTT	156	

Table 10: LMD-9 St1Cas9 guide target sequences (Examples 5-7)

Gene	ID	bp	Target		PAM	3'
			Sequence	SEQ ID NO:		
EMX1	G1	19	GAGGCAGGCAGGCTCTCCGA	157	GGAGAA	GGCCAA
	G1	23	GCCCAGGCAGGCAGGCTCTCCGA	158	GGAGAA	GGCCAA
	G2	20	GAGGAAGGGCCTGAGTCCGA	159	GCAGAA	GAAGAA
	G3	19	GGAGGACAAAGTACAAACG	160	GCAGAA	GCTGGA
	G4	20	GAAGGGCCTGAGTCCGAGCA	161	GAAGAA	GAAGGG
	G5	20	GGGCCTGAGTCCGAGCAGAA	162	GAAGAA	GGGCTC
	G6	19	GAGGGAGGGAGGGGCACAGA	163	TGAGAA	ACTCAG
	G6	22	GCCAGGGAGGGAGGGGCACAGA	164	TGAGAA	ACTCAG
	G7	19	GCAAACGGCAGAAGCTGGAG	165	GAGGAA	GGGCCT
	G7	22	GTACAAACGGCAGAAGCTGGAG	166	GAGGAA	GGGCCT
	G8	21	GTACAAACGGCAGAAGCTGGA	167	GGAGGA	AGGGCC
FANCF	G1	20	GGAAAAGCGATCCAGGTGCT	168	GCAGAA	GGGATT
	G2	20	GTAGGTAGTGCTTGAGACCG	169	CCAGAA	GCTCGG
	G4	20	GTAGGGCCTTCGCGCACCTC	170	ATGGAA	TCCCTT
	G5	19	GCTTGAGACCGCCAGAAGC	171	TCGGAA	AAGCGA
	G5	21	GTGCTTGAGACCGCCAGAAGC	172	TCGGAA	AAGCGA
	G6	20	GGCCTGGAAGTTCGCTAATC	173	CCGGAA	CTGGAC

	G7	20	GGCGACTCTCTGCGTACTGA	174	TTGGAA	CATCCG
RUNX1	G1	19	GCTAATCAGATGGAAGCTCT	175	TCAGAA	ATGTTT
	G1	21	GTACTAATCAGATGGAAGCTCT	176	TCAGAA	ATGTTT
	G2	19	GATCATTGAGTCCCCCGCCT	177	TCAGAA	GTGGGT
	G2	22	GAAATCATTGAGTCCCCCGCCT	178	TCAGAA	GTGGGT
	G3	19	GGTTTTCGCTCCGAAGGTA	179	AAAGAA	ATCATT
	G4	20	GAGGTATCCAGCAGAGGGGA	180	GAAGAA	AGAGAG
	G5	19	GCAGAGGTATCCAGCAGAGG	181	GGAGAA	GAAAGA
	G5	23	GTCCCAGAGGTATCCAGCAGAGG	182	GGAGAA	GAAAGA
	G7	22	GAATTCCTCTCACAAACAAGAC	183	AGGGAA	CTGGCA
ATP1A1	G1	20	GCAGCTTGGATGCTATAAGC	184	CAAGAA	ACAAAG
	G2	21	GCTTATAGCATCCAAGCTGCT	185	ACAGAA	GAGGAA
	G3	19	GCAAATCCATATGCTGAATT	186	ACAGAA	CTCACA
	G3	20	GACAAATCCATATGCTGAATT	187	ACAGAA	CTCACA
	G3	25	GTACTACAAATCCATATGCTGAATT	188	ACAGAA	CTCACA
	G7	20	GCATCCAAGCTGCTACAGAA	189	GAGGAA	CCTCAA
	G8	19	GCATCCAAGCTGCTACAGA	190	AGAGGA	ACCTCA
VEGF	G2	20	GAGAGGGACACACAGATCTA	191	TTGGAA	TCCTGG
	G3	20	GAGAGCCGTTCCCTCTTTGC	192	TAGGAA	TATTGA
	G4	20	GGGCATTGGCGAGGAGGGAG	193	CAGGAA	AGTGAG
	G5	19	GAAAATTACCCATCCGCCC	194	CCGGAA	ACTCTG
	G5	23	GCCTGAAAATTACCCATCCGCCC	195	CCGGAA	ACTCTG
Hpd	G1	20	GGTACCCCACGCAGAAAGCT	196	CGAGAA	CGGGGC
	G2	20	GTTGGGGCCTCGAATCCAGG	197	TAAGAA	ACGGCC

	G3	20	GATACCACACACACCCTGGT	198	GGAGAA	GATCAA
	G4	21	GTAGTAAGAAGATGGGGCGGC	199	CAAGAA	CTCCGT
	G5	21	GCATATCCTAGTCGACTATGA	200	CGAGAA	AGGCTA
Pcsk9	G1	19	GGTCACTGCTCATCTTCAC	201	CAAGAA	GCCAGG
	G2	19	GGGTCACTGCTCATCTTCAC	202	CAAGAA	GCCAGG
	G3	20	GAGGTCACTGCTCATCTTCAC	203	CAAGAA	GCCAGG
	G4	19	GCACGACGCCTCCCGCTCCT	204	GGAGAA	GCTGGA
	G5	20	GCCACGACGCCTCCCGCTCCT	205	GGAGAA	GCTGGA
	G6	20	GAGACCCATGTCCACTGCCA	206	CCAGAA	GGACCA
Pck1	G1	20	GGTGGGAACTCACTACTCGG	207	GAAGAA	ATGCTT
	G2	20	GGCATAACTAACCCCGAAGG	208	CAAGAA	GAAATA
	G3	20	GGGCACTGGCTGGCAGGGGT	209	GCAGAA	TCTCGA
	G4	20	GTATTTGCCGAAGTTGTAGCC	210	GAAGAA	GGGTCG
FANCF B	G1	21	GCAAGCGCTCCCACAGGCTGC	211	TGAGAA	ACCTGG
	G2	19	GCCTGTGGGAGCGCTTGCC	212	TCAGAA	CAACTT
	G3	19	GCCTTTGTCTCGTCGGCCC	213	CAAGAA	GAGTTG
	G4	19	GCAAAGACTTCCGAATTCC	214	CCAGAA	GCCAGT
	G5	22	GTCAACGTTTGCACTATGACCT	215	TCAGAA	AGGCAT
	G6	22	GCTTTACAGGTCTCCAGGGCAG	216	TTAGAA	CTTTAT
	G7	22	GTAATAACACAGCATTGCCTAT	217	ACAGAA	CTGAGG
	G8	19	GCTGTGTTATTACTTGAAT	218	ATAGAA	TATATA
	G9	23	GACACACGAAGGCATATATTTGG	219	TGAGAA	CATTGT
	G10	22	GTCTCGTCGGCCCCAAGAAGAG	220	TTGGAA	CCCGGC
	G11	22	GACCTTCAGAAAGGCATTTGGG	221	TTGGAA	CTGAGT

AAVS1	G2	20	GAGGGGACAGATAAAAGTAC	222	CCAGAA	CCAGAG
	G4'	23	GAAATGGGGGTGTGTCACCAGAT	223	AAGGAA	TCTGCC
	G5'	22	GTTAGACCCAATATCAGGAGAC	224	TAGGAA	GGAGGA
	G6'	19	GAGCCACATTAACCGGCCC	225	TGGGAA	TATAAG
	G8'	22	GACTAGCTGAGCTCTCGGACCC	226	CTGGAA	GATGCC
	G9'	20	GAAGATGCCATGACAGGGGG	227	CTGGAA	GAGCTA
CFTR	G1	20	GCTATTTTTATGGGACATTT	228	TCAGAA	CTCCAA
	G3	20	GGAGAGTTTGGGGAAAAAAG	229	GAAGAA	TTCTAT
	G4	23	GTATAGAGTTGATTGGATTGAGA	230	ATAGAA	TTCTTC
	G5	20	GCCTTCTCTCTAAAGGCTCA	231	TCAGAA	TCCTCT
	G6	20	GCAGTATCGCCTCTCCCTGC	232	TCAGAA	TCTGGT
	G7	21	GACTGGAGAGTTTGGGGAAAA	233	AAGGAA	GAATTC

Table 11: LMG 18311 St1Cas9 guide target sequences (Examples 5-7)

Gene	ID	bp	Target		PAM	3'
			Sequence	SEQ ID NO:		
FANCF	G202	20	GGCGGCTGCACAACCAGTGG	234	AGGCAA	GAGGGC
RUNX1	G201	23	GAAACAAGCTGCCATTTCATTAC	235	AGGCAA	AGCTGA
	G202	23	GCCATTTCATTACAGGCAAAGCT	236	GAGCAA	AAGTAG
	205	20	GAGGTGAGTACATGCTGGTC	237	TTGTAA	TATCTA
AAVS1	G201	20	GGACACAGGATCCCTGGAGG	238	CAGCAA	ACATGC
Grin2B	G206	20	GGCTTCCTGGTCTGTGTCAT	239	CAGCAA	ACACCA

ATP1A1	G202	23	GTTCATAACCATTAAGTAATGAG	240	TGGTAA	TTGAGA
	G203	20	GTAATGAGTTCATAACCATT	241	AAGTAA	TGAGTG

Table 12: CNRZ 1066 St1Cas9 guide target sequences (Examples 5-7)

Gene	ID	bp	Target		PAM	3'
			Sequence	SEQ ID NO:		
EMX1	G101	20	GGTGTGGTTCCAGAACCGGA	242	GGACAA	AGTACA
	G102	21	GTTCCAGAACCGGAGGACAAA	243	GTACAA	ACGGCA
ATP1A1	G102	21	GCTTGGATGCTATAAGCCAAG	243	AAACAA	AGAATC
	G103	21	GGAGAAGATATCTGATGTGTA	245	CTACAA	ATCCAT
	G104	20	GGTAATTGAGAAGAAGTGGG	246	AGACAA	AGACGG

Table 13: TH1477 St1Cas9 guide target sequences (Examples 5-7)

Gene	ID	bp	Target		PAM	3'
			Sequence	SEQ ID NO:		
FANCF	303a	19	GAGAGTCGCCGTCTCCAAG	247	GTGAAA	GCGGAA
	304	20	GCTTGAGACCGCCAGAAGCT	248	CGGAAA	AGCGAT
	308	21	GCTCTTCGTAGTGGTGCATTT	249	AGGAAA	AGACAA
	309	19	GAATATATAGTTTACAAAA	250	ATGAAA	ATTACA
RUNX1	302	21	GTCTGAAGCCATCGCTTCCTC	251	CTGAAA	ATGCAC
	304	19	GATTTCTTTTACCTTCGGA	252	GCGAAA	ACCAAG
Grin2b	301	21	GTTCAAGGATTTCTGAGGCTT	253	TTGAAA	GTTTCA
	303	20	GTATTTGCTCTGCAGAATGA	254	GAGAAA	ATGAAA

	304	20	GGAGTTGGGTTTGGTGCTCA	255	ATGAAA	GGAGAT
	305	20	GTCGACTCCCTGCAAACACA	256	AAGAAA	GAGCAT
	306	21	GTGGCCATCAAGGATGCCCAC	257	GAGAAA	GATGAT
	307	20	GTTAAAATAGGATCTACATC	258	ACGTAA	
AAVS1	301	23	GCCACTAGGGACAGGATTGGTGA	259	CAGAAA	AGCCCC

Table 14: Sequences described herein

SEQ ID NO(s):	Description
1	sgRNA, FIG. 1b (positions 21-103)
2-3	target nucleic acid sequence, FIG. 1c, sense and antisense; FANCF
4-5	target nucleic acid sequence, FIG. 3c, sense and antisense
6	amino acid sequence; FIG. 3c
7-11	sgRNAs, FIG. 5c, in order shown
12-13	target nucleic acid sequence, FIG. 6a, sense and antisense
14	amino acid sequence; FIG. 6a
15	SaCas9 amino acid sequence, FIG. 9
16	St1Cas9_LMD-9 amino acid sequence, FIGs. 9 and 10
17	St1Cas9_LMG18311 amino acid sequence, FIGs. 9 and 10
18	St1Cas9_CNRZ1066 amino acid sequence, FIGs. 9 and 10
19	St1Cas9 TH1477 amino acid sequence
20-21	Guide sequence, sense and antisense, FIG. 11
22	Full nucleic acid sequence, FIG. 16
23-24	SV40 NLS nucleic acid sequences, FIGs. 16, 23a, 23b, 25a, 25b, 27a, 27b
25	ST1Cas9 nucleic acid sequence (LMD-9), FIG. 16
26-27	Linker between NLS and ST1Cas9, nucleic acid sequences, FIGs. 16, 23a, 23b, 25a, 25b, 27a, 27b
28	Full amino acid sequence, FIG. 17
29	SV40 NLS amino acid sequence, FIGs. 17, 24, 26, 28
30	ST1Cas9 amino acid sequence (LMD-9), FIG. 17
31-32	Linker between NLS and ST1Cas9, amino acid sequences, FIGs. 17, 24, 26, 28
33	Full nucleic acid sequence, FIGs. 23a-23b
34	ST1Cas9 hybrid nucleic acid sequence (LMD-9/LMG18311), FIGs. 23a-23b
35	Full amino acid sequence, FIG. 24
36	ST1Cas9 hybrid amino acid sequence (LMD-9/LMG18311), FIG. 24
37	Full nucleic acid sequence, FIGs. 25a-25b
38	ST1Cas9 hybrid nucleic acid sequence (LMD-9/CNRZ1066), FIGs. 25a-25b
39	Full amino acid sequence, FIG. 26
40	ST1Cas9 hybrid amino acid sequence (LMD-9/CNRZ1066), FIG. 26
41	Full nucleic acid sequence, FIGs. 27a-27b
42	ST1Cas9 hybrid nucleic acid sequence (LMD-9/TH1477), FIGs. 27a-27b
43	Full amino acid sequence, FIG. 28
44	ST1Cas9 hybrid amino acid sequence (LMD-9/TH1477), FIG. 28
45	Full nucleic acid sequence, FIGs. 29a-29b
46	rAPOBEC1 nucleic acid sequence; FIGs. 29a-29b
47	UGI nucleic acid sequence, FIGs. 29a-29b
48	3xHA nucleic acid sequence, FIGs. 29a-29b
49	Full amino acid sequence, FIG. 30
50	rAPOBEC1 amino acid sequence; FIGs. 30, 32, 34
51	UGI amino acid sequence, FIGs. 30, 32, 34
52	3xHA amino acid sequence, FIGs. 30, 32, 34
53	Full nucleic acid sequence, FIGs. 31a-31b
54	ST1Cas9 nucleic acid sequence (LMD-9/LMG18311 hybrid), FIGs. 31a-31b
55	Full amino acid sequence, FIG. 32
56	ST1Cas9 amino acid sequence (LMD-9/LMG18311 hybrid), FIG. 32
57	Full nucleic acid sequence, FIGs. 33a-33b
58	ST1Cas9 nucleic acid sequence (LMD-9/CNRZ1066 hybrid), FIGs. 33a-33b
59	Full amino acid sequence, FIG. 34
60	ST1Cas9 amino acid sequence (LMD-9/CNRZ1066 hybrid), FIG. 34

61	Nucleoplasmin NLS amino acid sequence
62-65	NLS amino acid sequences
66-67	target nucleic acid sequence, FIG. 1c, sense and antisense; EMX1
68-69	target nucleic acid sequence, FIG. 1c, sense and antisense; RUNX1
70-98	guide target sequences, Table 3
99-108	guide target sequences, Table 4
109-113	guide target sequences, Table 5
114-118	guide target sequences, Table 6
119-124	sgRNAs (DNA and RNA sequences), Table 7
125-126	TBG and LP1B promoter sequences, Table 8
127-156	PCR Primer sequences, Table 9
157-259	guide target sequences, Tables 10-13
260-263	amino acid sequences of ST1Cas9 C-terminal region (LMD-9, LMG18311, CNRZ1066, and TH1477), FIG. 35
264-267	amino acid sequences of ST1Cas9 N-terminal region (LMD-9, LMG18311, CNRZ1066, and TH1477), FIG. 10

While the invention has been described in connection with specific embodiments thereof, it will be understood that the scope of the claims should not be limited by the preferred embodiments set forth in the examples, but should be given the broadest interpretation consistent with the description as a whole. In the claims, the word "comprising" is used as an open-ended term, substantially equivalent to the phrase "including, but not limited to". The singular forms "a", "an" and "the" include corresponding plural references unless the context clearly dictates otherwise.

REFERENCES

1. Komor, A.C., Badran, A.H. & Liu, D.R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **169**, 559 (2017).

2. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

3. Koonin, E.V., Makarova, K.S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67-78 (2017).

4. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182 (2017).

5. Hille, F. et al. The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239-1259 (2018).

6. Esvelt, K.M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10**, 1116-1121 (2013).

7. Ran, F.A. et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191 (2015).

8. Chen, F. et al. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun* **8**, 14958 (2017).

9. Anderson, E.M. et al. Lactobacillus gasseri CRISPR-Cas9 characterization In Vitro reveals a flexible mode of protospacer-adjacent motif recognition. *PLoS One* **13**, e0192181 (2018).

10. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771 (2015).

11. Colella, P., Ronzitti, G. & Mingozzi, F. Emerging Issues in AAV-Mediated In Vivo Gene Therapy. *Mol Ther Methods Clin Dev* **8**, 87-104 (2018).

12. Friedland, A.E. et al. Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol* **16**, 257 (2015).

13. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**, 14500 (2017).

14. Ibraheim, R. et al. All-in-One Adeno-associated Virus Delivery and Genome Editing by Neisseria meningitidis Cas9 in vivo. *bioRxiv* (2018).

15. Mir, A., Edraki, A., Lee, J. & Sontheimer, E.J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* **13**, 357-365 (2018).

16. Anguela, X.M. et al. Robust ZFN-mediated genome editing in adult hemophilic mice. *Blood* **122**, 3283-3287 (2013).

17. Li, H. et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217-221 (2011).

18. Sharma, R. et al. In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777-1784 (2015).

19 Nami, F. et al. Strategies for In Vivo Genome Editing in Nondividing Cells. *Trends Biotechnol* (2018).

20. He Z, Proudfoot C, Mileham A, J., McLaren DG, Whitelaw BA, Lillico SG. Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnology and Bioengineering.* **112**(5): 1060-4 (2015).

21. Byrne SM, Ortiz L, Mali P, Aach J, Church GM. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.* **43**(3):e21 (2015).

22. Schneller, J.L., Lee, C.M., Bao, G. & Venditti, C.P. Genome editing for inborn errors of metabolism: advancing towards the clinic. *BMC Med* **15**, 43 (2017).

23. Lau, C.H. & Suh, Y. In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease. *F1000Res* **6**, 2153 (2017).

24. Bolotin, A. et al. Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. *Nat Biotechnol* **22**, 1554-1558 (2004).

25. Barrangou, R. & Horvath, P. A decade of discovery: CRISPR functions and applications. *Nat Microbiol* **2**, 17092 (2017).

26. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561 (2005).
27. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
28. Horvath, P. et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *J Bacteriol* **190**, 1401-1412 (2008).
29. Deveau, H. et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**, 1390-1400 (2008).
30. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71 (2010).
31. Briner, A.E. et al. Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **56**, 333-339 (2014).
32. Hynes, A.P. et al. An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. *Nat Microbiol* **2**, 1374-1380 (2017).
33. Chari, R., Mali, P., Moosburner, M. & Church, G.M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* **12**, 823-826 (2015).
34. Kleinstiver, B.P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485 (2015).
35. Agudelo, D. et al. Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat Methods* **14**, 615-620 (2017).
36. Dalvai, M. et al. A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep* **13**, 621-633 (2015).
37. Guschin, D.Y. et al. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**, 247-256 (2010).
38. Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168 (2014).
39. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148 (2016).
40. Tomoeda, K. et al. Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type III and hawkinsinuria. *Mol Genet Metab* **71**, 506-510 (2000).
41. Ruetschi, U. et al. Mutations in the 4-hydroxyphenylpyruvate dioxygenase gene (HPD) in patients with tyrosinemia type III. *Hum Genet* **106**, 654-662 (2000).
42. Russell, S. et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390**, 849-860 (2017).
43. George, L.A. et al. Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N Engl J Med* **377**, 2215-2227 (2017).
44. Nathwani, A.C. et al. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med* **371**, 1994-2004 (2014).
45. McKay, T.R. et al. Perinatal gene transfer to the liver. *Curr Pharm Des* **17**, 2528-2541 (2011).
46. Wang, L. et al. AAV8-mediated hepatic gene transfer in infant rhesus monkeys (Macaca mulatta). *Mol Ther* **19**, 2012-2020 (2011).
47. Wang, L., Wang, H., Bell, P., McMenamin, D. & Wilson, J.M. Hepatic gene transfer in neonatal mice by adeno-associated virus serotype 8 vector. *Hum Gene Ther* **23**, 533-539 (2012).
48. Yang, Y. et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat Biotechnol* **34**, 334-338 (2016).
49. Morrow, G. & Tanguay, R.M. Biochemical and Clinical Aspects of Hereditary Tyrosinemia Type 1. *Adv Exp Med Biol* **959**, 9-21 (2017).

50. Grompe, M. Fah Knockout Animals as Models for Therapeutic Liver Repopulation. *Adv Exp Med Biol* **959**, 215-230 (2017).
51. Endo, F. et al. Complete rescue of lethal albino c14CoS mice by null mutation of 4-hydroxyphenylpyruvate dioxygenase and induction of apoptosis of hepatocytes in these mice by in vivo retrieval of the tyrosine catabolic pathway. *J Biol Chem* **272**, 24426-24432 (1997).
52. Pankowicz, F.P. et al. Reprogramming metabolic pathways in vivo with CRISPR/Cas9 genome editing to treat hereditary tyrosinaemia. *Nat Commun* **7**, 12642 (2016).
53. Nathwani, A.C. et al. Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver. *Blood* **107**, 2653-2661 (2006).
54. McIntosh, J. et al. Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human factor VIII variant. *Blood* **121**, 3335-3344 (2013).
55. Pawluk, A., Davidson, A.R. & Maxwell, K.L. Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol* **16**, 12-17 (2018).
56. Shin, J. et al. Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci Adv* **3**, e1701620 (2017).
57. Rauch, B.J. et al. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* **168**, 150-158 e110 (2017).
58. Rousseau, B.A., Hou, Z., Gramelspacher, M.J. & Zhang, Y. Programmable RNA Cleavage and Recognition by a Natural CRISPR-Cas9 System from Neisseria meningitidis. *Mol Cell* **69**, 906-914 e904 (2018).
59. Strutt, S.C., Torrez, R.M., Kaya, E., Negrete, O.A. & Doudna, J.A. RNA-dependent RNA targeting by CRISPR-Cas9. *Elife* **7** (2018).
60. Zhang, Y., Rajan, R., Seifert, H.S., Mondragon, A. & Sontheimer, E.J. DNase H Activity of Neisseria meningitidis Cas9. *Mol Cell* **60**, 242-255 (2015).
61. Stephenson, A.A., Raper, A.T. & Suo, Z. Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9. *J Am Chem Soc* **140**, 3743-3750 (2018).
62. Ma, E., Harrington, L.B., O'Connell, M.R., Zhou, K. & Doudna, J.A. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol Cell* **60**, 398-407 (2015).
63. Dugar, G. et al. CRISPR RNA-Dependent Binding and Cleavage of Endogenous RNAs by the Campylobacter jejuni Cas9. *Mol Cell* **69**, 893-905 e897 (2018).
64. Chen, J.S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* (2018).
65. Boettcher, M. et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat Biotechnol* **36**, 170-178 (2018).
66. Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res* **42**, 2577-2590 (2014).
67. Goudy, K.S., Annoni, A., Naldini, L. & Roncarolo, M.G. Manipulating Immune Tolerance with Micro-RNA Regulated Gene Therapy. *Frontiers in microbiology* **2**, 221 (2011).
68. Charlesworth, C.T. et al. Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. *bioRxiv* (2018).
69. Moreno, A.M. et al. Exploring protein orthogonality in immune space: a case study with AAV and Cas9 orthologs. *bioRxiv* (2018).
70. Wagner, D.L. et al. High prevalence of S. pyogenes Cas9-specific T cell sensitization within the adult human population - A balanced effector/regulatory T cell response. *bioRxiv* (2018).
71. Chew, W.L. et al. A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods* **13**, 868-874 (2016).
72. Muller, M. et al. Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Mol Ther* **24**, 636-644 (2016).
73. Karvelis, T., Gasiunas, G. & Siksnys, V. Methods for decoding Cas9 protospacer adjacent motif (PAM) sequences: A brief overview. *Methods* **121-122**, 3-8 (2017).

74. Leenay, R.T. & Beisel, C.L. Deciphering, Communicating, and Engineering the CRISPR PAM. *J Mol Biol* **429**, 177-191 (2017).

75. Rock, J.M. et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol* **2**, 16274 (2017).

76. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A* **103**, 15611-15616 (2006).

77. Chen, H., Choi, J. & Bailey, S. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *J Biol Chem* **289**, 13284-13294 (2014).

78. Hu, J.H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63 (2018).

79. Kleinstiver, B.P. et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293-1298 (2015).

80. Chatterjee, P., Jakimo, N. & Jacobson, J.M. Divergent PAM Specificity of a Highly-Similar SpCas9 Ortholog. *bioRxiv* (2018).

81. Gray, S.J. et al. Production of recombinant adeno-associated viral vectors and use in in vitro and in vivo administration. *Curr Protoc Neurosci* **Chapter 4**, Unit 4 17 (2011).

82. Aurnhammer, C. et al. Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. *Hum Gene Ther Methods* **23**, 18-28 (2012).

83. Grompe, M. et al. Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. *Genes Dev* **7**, 2298-2307 (1993).

84. Yardeni, T., Eckhaus, M., Morris, H.D., Huizing, M. & Hoogstraten-Miller, S. Retro-orbital injections in mice. *Lab Anim (NY)* **40**, 155-160 (2011).

85. Cyr, D., Giguere, R., Villain, G., Lemieux, B. & Drouin, R. A GC/MS validated method for the nanomolar range determination of succinylacetone in amniotic fluid and plasma: an analytical tool for tyrosinemia type I. *J Chromatogr B Analyt Technol Biomed Life Sci* **832**, 24-29 (2006).

86. Chatterjee, P., Jakimo, N. & Jacobson, J.M. Minimal PAM specificity of a highly similar SpCas9 ortholog. Sci Adv 4, eaau0766 (2018).

WHAT IS CLAIMED IS:

1. A sgRNA for modification of a target polynucleotide in a cell, comprising:

(a) a guide segment comprising a guide sequence corresponding to a region of the target polynucleotide;

(b) a first hairpin-forming segment located 3' to the guide sequence, the first hairpin hairpin-forming segment being capable of forming a hairpin comprising a stem portion and a loop portion, wherein the stem portion does not comprise a sequence corresponding to an RNA polymerase III termination signal.

2. The sgRNA of claim 1, wherein the stem portion does not comprise more than 4 consecutive uracil nucleotides.

3. The sgRNA of claim 1, wherein the stem portion does not comprise more than 3 consecutive uracil nucleotides.

4. The sgRNA of claim 1, wherein the stem portion comprises a first stem portion and a second stem portion, wherein the first stem portion does not comprise a sequence corresponding to an RNA polymerase III termination signal.

5. The sgRNA of claim 4, wherein the first stem portion does not comprise more than 4 consecutive uracil nucleotides.

6. The sgRNA of claim 5, wherein the first stem portion does not comprise more than 3 consecutive uracil nucleotides.

7. The sgRNA of any one of claims 4 to 6, wherein the first stem portion and second stem portion are separated by a first bulge portion.

8. The sgRNA of any one of claims 1 to 8, wherein the loop portion comprises or consists of a sequence of 3 to 6 nucleotides.

9. The sgRNA of claim 8, wherein the loop portion comprises or consists of a sequence of 3 to 5 nucleotides.

10. The sgRNA of claim 9, wherein the loop portion comprises or consists of a sequence of 4 nucleotides.

11. The sgRNA of claim 10, wherein the loop portion comprises or consists of the nucleotide sequence $N^1N^2N^3N^4$, wherein N^1, N^2, and N^3 are each independently A, C, G or U, and N^4 is C or G.

12. The sgRNA of claim 10 or 11, wherein the loop portion comprises or consists of the nucleotide sequence $N^1N^2N^3N^4$, wherein N^1, N^3, and N^4 are each independently A, C, G or U, and N^2 is U, G or A.

13. The sgRNA of claim 11 or 12, wherein N^1 is G.

14. The sgRNA of any one of claims 11 to 13, wherein N^2 is U.

15. The sgRNA of any one of claims 11 to 14, wherein N^3 is A.

16. The sgRNA of any one of claims 11 to 15, wherein N^4 is C.

17. The sgRNA of any one of claims 10 to 16, wherein the loop portion comprises or consists of the nucleotide sequence GUAC.

18. The sgRNA of any one of claims 4 to 17, wherein the second stem portion comprises or consists of a hybrid of 4 nucleotide pairs.

19. The sgRNA of claim 18, wherein the fourth pair of the hybrid of the second stem portion, distal to the first stem portion, is a G-C pair.

20. The sgRNA of claim 18 or 19, wherein the hybrid of the second stem portion comprises or consists of the sequence 5'-UCUG-3' hybridized to the sequence 5'-CAGA-3'.

21. The sgRNA of any one of claims 4 to 20, wherein the first stem portion comprises or consists of a hybrid of at least 5 nucleotide pairs.

22. The sgRNA of any one of claims 4 to 21, wherein the first stem portion comprises or consists of a hybrid of not more than 12 nucleotide pairs.

23. The sgRNA of claim 21 or 22, wherein the first stem portion comprises or consists of a hybrid of 6 to 10 nucleotide pairs.

24. The sgRNA of claim 23, wherein the first stem portion comprises or consists of a hybrid of 7 to 9 nucleotide pairs.

25. The sgRNA of claim 24, wherein the first stem portion comprises or consists of a hybrid of 8 nucleotide pairs.

26. The sgRNA of any one of claims 4 to 25, wherein the first stem portion does not comprise a mismatch.

27. The sgRNA of any one of claims 4 to 26, wherein the hybrid of the first stem portion comprises or consists of the sequence 5'-UCUUUGUA-3' hybridized to the sequence 5'-UACAAAGA-3'.

28. The sgRNA of any one of claims 4 to 24, wherein the first stem portion comprises a single mismatch.

29. The sgRNA of claim 28, wherein the hybrid of the first stem portion comprises or consists of the sequence 5'-GUCUUUGUA-3' hybridized to the sequence 5'-UACAAAGAU-3'.

30. The sgRNA of any one of claims 1 to 29, further comprising one or more additional hairpin-forming segments located 3' to the first hairpin-forming segment.

31. The sgRNA of claim 30, further comprising one or more linker segments located between the first hairpin-forming segment and additional hairpin-forming segments, and/or between the additional hairpin-forming segments.

32. A nucleic acid comprising a nucleotide sequence encoding the sgRNA of any one of claims 1 to 31.

33. A vector comprising the nucleic acid of claim 32.

34. The vector of claim 33, further comprising a nucleotide sequence encoding a CRISPR nuclease.

35. The vector of claim 34, wherein the CRISPR nuclease is a Cas9 enzyme.

36. The vector of claim 34 or 35, wherein the CRISPR nuclease is derived from non-pathogenic bacteria.

37. The vector of any one of claims 34 to 36, wherein the CRISPR nuclease is a *Streptococcus thermophilus* Cas9 nuclease.

38. The vector of any one of claims 34 to 37, wherein the CRISPR nuclease is a type II Cas9 nuclease.

39. The vector of any one of claims 34 to 38, wherein the CRISPR nuclease is a *Streptococcus thermophilus* type II-A CRISPR1-Cas9 (St1Cas9).

40. The vector of any one of claims 34 to 39, wherein the CRISPR nuclease further comprises one or more nuclear localization signal (NLS) and the vector further comprises one or more nucleotide sequences encoding the one or more NLSs.

41. The vector of claim 40, wherein the CRISPR nuclease comprises a first NLS at its amino terminal end and a second NLS at its carboxy terminal end, and the vector comprises NLS-encoding nucleotide sequences flanking the CRISPR nuclease-encoding nucleotide sequence.

42. The vector of claim 33, further comprising a promoter operably-linked to the nucleotide sequence encoding the sgRNA.

43. The vector of any one of claims 34 to 41, further comprising one or more promoters operably-linked to the nucleotide sequence encoding the sgRNA and/or the nucleotide sequence encoding the CRISPR nuclease.

44. The vector of claim 43, wherein the nucleotide sequence encoding the sgRNA and or the nucleotide sequence encoding the CRISPR nuclease are both operably linked to a single promoter.

45. The vector of claim 43, wherein the nucleotide sequence encoding the sgRNA is operably linked to a first promoter and the nucleotide sequence encoding the CRISPR nuclease is operably linked to a second promoter, wherein the first and second promoters may be the same or different.

46. The vector of claim 45, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in the same orientation within the vector.

47. The vector of claim 45, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in opposite orientations within the vector.

48. The vector of any one of claims 33 to 47, wherein the vector is a viral vector.

49. The vector of claim 48, wherein the vector is an adeno-associated virus (AAV) vector.

50. A host cell comprising the nucleic acid of claim 32 or the vector of any one of claims 33 to 49.

51. A composition comprising the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, or the host cell of claim 50.

52. The composition of claim 51, further comprising a pharmaceutically acceptable carrier.

53. A system comprising the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the host cell of claim 50, and/or the composition of claim 51 or 52.

54. A system comprising the vector of claim 33 and a further vector comprising a nucleotide sequence encoding a CRISPR nuclease.

55. The system of claim 54, wherein the CRISPR nuclease is as defined in any one of claims 35 to 41.

56. The system of claim 54 or 55, wherein the vector of claim 33 further comprises a promoter operably-linked to the nucleotide sequence encoding the sgRNA and further vector further comprises a promoter operably-linked to the nucleotide sequence encoding the CRISPR nuclease.

57. A method of modifying a target polynucleotide in a cell, comprising contacting the cell with the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56.

58. The method of claim 57, which is an *in vitro* method.

59. The method of claim 57, which is an *in vivo* method and the cell is in a subject.

60. Use of the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56, for modifying a target polynucleotide in a cell.

61. Use of the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56, for the preparation of a medicament for modifying a target polynucleotide in a cell.

62. The sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56, for use in modifying a target polynucleotide in a cell.

63. A method of preventing or treating a condition associated with a target polynucleotide in a subject in need thereof, comprising administering to the subject an effective amount the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56.

64. The method of claim 63, wherein the condition is a metabolic condition.

65. The method of claim 63 or 64, wherein the condition is a hepatic condition.

66. Use of the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56, for preventing or treating a condition associated with a target polynucleotide in a subject.

67. Use of the sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56, for the preparation of a medicament for preventing or treating a condition associated with a target polynucleotide in a subject.

68. The use of claim 66 or 67, wherein the condition is a metabolic condition.

69. The use of any one of claims 66 to 68, wherein the condition is a hepatic condition.

70. The sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56, for use in preventing or treating a condition associated with a target polynucleotide in a subject.

71. The sgRNA, nucleic acid, vector, composition and/or system for use of claim 70, wherein the condition is a metabolic condition.

72. The sgRNA, nucleic acid, vector, composition and/or system for use of claim 70 or 71, wherein the condition is a hepatic condition.

73. The sgRNA of any one of claims 1 to 31, the nucleic acid of claim 32, the vector of any one of claims 33 to 49, the host cell of claim 50, the composition of claim 51 or 52 and/or the system of any one of claims 53 to 56, for use as a medicament.

74. An isolated CRISPR nuclease polypeptide comprising a first domain and a second domain C-terminal to the first domain, wherein the first domain comprises a guide RNA-binding domain and a nuclease domain, and the second domain comprises a WED domain and a PAM-interacting domain, wherein the first and second domains are derived from different bacterial strains.

75. The isolated polypeptide of claim 74, wherein the first and second domains are derived from non-pathogenic bacteria.

76. The isolated polypeptide of claim 74 or 75, wherein the first and second domains are derived from different bacterial species.

77. The isolated polypeptide of claim 74 or 75, wherein the first and second domains are derived from different strains of the same bacterial species.

78. The isolated polypeptide of claim 77, wherein the first and second domains are derived from different strains of *Streptococcus thermophilus*.

79. The isolated polypeptide of any one of claims 74 to 78, wherein the CRISPR nuclease is a type II Cas9 nuclease.

80. The isolated polypeptide of any one of claims 74 to 79, wherein the CRISPR nuclease is a *Streptococcus thermophilus* type II-A CRISPR1-Cas9 (St1Cas9).

81. The isolated polypeptide of any one of claims 74 to 80, further comprising one or more nuclear localization signal (NLS).

82. The isolated polypeptide of claim 81, comprising a first NLS N-terminal to the first domain and a second NLS C-terminal to the second domain.

83. The isolated polypeptide of any one of claims 74 to 82, further comprising a cytidine deaminase domain or an adenosine deaminase domain.

84. The isolated polypeptide of claim 83, comprising a cytidine deaminase domain.

85. The isolated polypeptide of claim 84, wherein the cytidine deaminase is an APOBEC cytidine deaminase.

86. The isolated polypeptide of claim 84, wherein the cytidine deaminase domain comprises the amino acid sequence of SEQ ID NO: 50, or a functional fragment thereof, or a functional variant thereof.

87. The isolated polypeptide of claim 84 or 85, further comprising a uracil DNA glycosylase inhibitor (UGI) domain.

88. The isolated polypeptide of claim 87, wherein the UGI domain comprises the amino acid sequence of SEQ ID NO: 51, or a functional fragment thereof, or a functional variant thereof.

89. The isolated polypeptide of any one of claims 74 to 88, wherein the first domain is derived from *Streptococcus thermophilus* LMD-9, LMG8311, CNRZ1066 or TH1477.

90. The isolated polypeptide of any one of claims 74 to 89, wherein the second domain is derived from *Streptococcus thermophilus* LMD-9, LMG8311, CNRZ1066 or TH1477.

91. The isolated polypeptide of any one of claims 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* LMD-9 and the second domain is derived from *Streptococcus thermophilus* LMG8311, CNRZ1066 or TH1477.

92. The isolated polypeptide of any one of claims 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* LMG8311 and the second domain is derived from *Streptococcus thermophilus* LMD-9, CNRZ1066 or TH1477.

93. The isolated polypeptide of any one of claims 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* CNRZ1066 and the second domain is derived from *Streptococcus thermophilus* LMG8311, LMD-9 or TH1477.

94. The isolated polypeptide of any one of claims 74 to 90, wherein the first domain is derived from *Streptococcus thermophilus* TH1477 and the second domain is derived from *Streptococcus thermophilus* LMG8311, CNRZ1066 or LMD-9.

95. The isolated polypeptide of any one of claims 74 to 88, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 264, 265, 266, or 267, or a functional fragment of any thereof, or a functional variant of any thereof.

96. The isolated polypeptide of any one of claims 74 to 88, wherein the second domain comprises the amino acid sequence of SEQ ID NO: 260, 261, 262, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

97. The isolated polypeptide of any one of claims 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 264, or a functional fragment of any thereof, or a functional variant of any thereof,

and the second domain comprises the amino acid sequence of SEQ ID NO: 261, 262, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

98. The isolated polypeptide of any one of claims 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 265, or a functional fragment of any thereof, or a functional variant of any thereof, and the second domain comprises the amino acid sequence of SEQ ID NO: 260, 262, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

99. The isolated polypeptide of any one of claims 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 266, or a functional fragment of any thereof, or a functional variant of any thereof, and the second domain comprises the amino acid sequence of SEQ ID NO: 260, 261, or 263, or a functional fragment of any thereof, or a functional variant of any thereof.

100. The isolated polypeptide of any one of claims 74 to 88, 95 and 96, wherein the first domain comprises the amino acid sequence of SEQ ID NO: 267, or a functional fragment of any thereof, or a functional variant of any thereof, and the second domain comprises the amino acid sequence of SEQ ID NO: 260, 261, or 262, or a functional fragment of any thereof, or a functional variant of any thereof.

101. The isolated polypeptide of any one of claims 74 to 100, the first domain is connected to the second domains via a linker region.

102. The isolated polypeptide of any one of claims 74 to 101, wherein the polypeptide is capable of binding a PAM that is different from the PAM bound by a CRISPR nuclease from which the first domain is derived.

103. The isolated polypeptide of any one of claims 74 to 102, wherein the polypeptide binds a PAM comprising the sequence NNAGAA, NNGGAA, NNACAA, NNGCAA, NNGAAA or NNAAAA.

104. A nucleic acid comprising a nucleotide sequence encoding the isolated polypeptide of any one of claims 74 to 103.

105. A vector comprising the nucleic acid of claim 104.

106. The vector of claim 105, further comprising a nucleotide sequence encoding an sgRNA.

107. The vector of claim 106, wherein the sgRNA is the sgRNA of any one of claims 1 to 31.

108. The vector of any one of claims 105 to 107, further comprising one or more promoters operably-linked to the nucleotide sequence encoding the polypeptide and/or the nucleotide sequence encoding the sgRNA.

109. The vector of claim 108, wherein the nucleotide sequence encoding the polypeptide and the nucleotide sequence encoding the sgRNA are both operably linked to a single promoter.

110. The vector of claim 108, wherein the nucleotide sequence encoding the sgRNA is operably linked to a first promoter and the nucleotide sequence encoding the polypeptide is operably linked to a second promoter, wherein the first and second promoters may be the same or different.

111. The vector of claim 110, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in the same orientation within the vector.

112. The vector of claim 110, wherein (i) the first promoter and the nucleotide sequence encoding the sgRNA and (ii) the second promoter and the nucleotide sequence encoding the CRISPR nuclease are in opposite orientations within the vector.

113. The vector of any one of claims 105 to 112, wherein the vector is a viral vector.

114. The vector of claim 113, wherein the vector is an adeno-associated virus (AAV) vector.

115. A host cell comprising the nucleic acid of claim 104 or the vector of any one of claims 105-113.

116. A composition comprising the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, or the host cell of claim 115.

117. The composition of claim 116, further comprising a pharmaceutically or biologically acceptable carrier.

118. A system comprising the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117.

119. A system comprising the vector of claim 105 and a further vector comprising a nucleotide sequence encoding an sgRNA.

120. A method of modifying a target polynucleotide in a cell, comprising contacting the cell with the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119.

121. The method of claim 120, which is an *in vitro* method.

122. The method of claim 120, which is an *in vivo* method and the cell is in a subject.

123. Use of the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119, for modifying a target polynucleotide in a cell.

124. Use of the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119, for the preparation of a medicament for modifying a target polynucleotide in a cell.

125. The polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119, for use in modifying a target polynucleotide in a cell.

126. A method of preventing or treating a condition associated with a target polynucleotide in a subject in need thereof, comprising administering to the subject an effective amount the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119.

127. The method of claim 126, wherein the condition is a metabolic condition.

128. The method of claim 126 or 127, wherein the condition is a hepatic condition.

129. Use of the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119, for preventing or treating a condition associated with a target polynucleotide in a subject.

130. Use of the polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119, for the preparation of a medicament for preventing or treating a condition associated with a target polynucleotide in a subject.

131. The use of claim 129 or 130, wherein the condition is a metabolic condition.

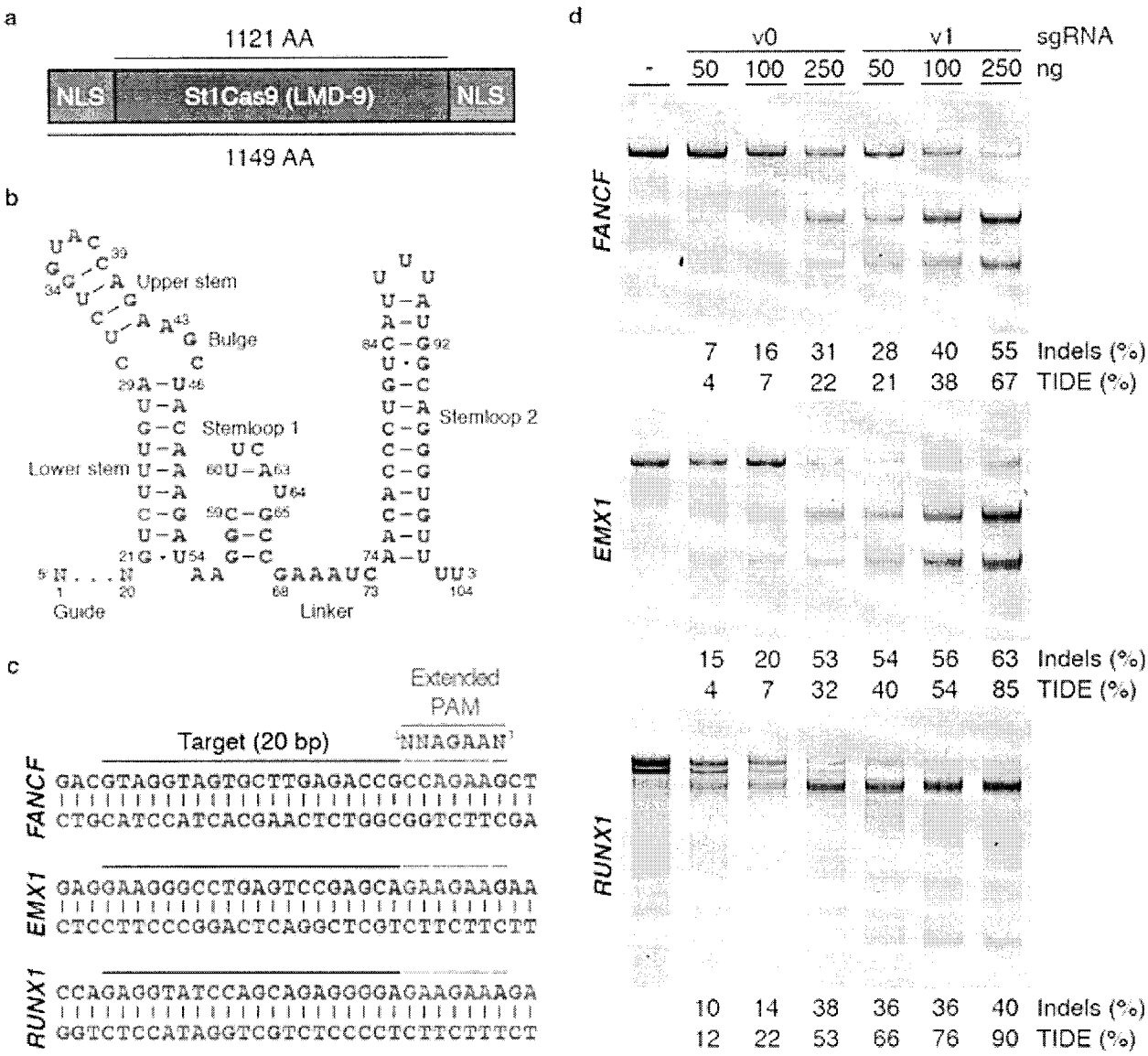
132. The use of any one of claims 129 to 131, wherein the condition is a hepatic condition.

133. The polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119, for use in preventing or treating a condition associated with a target polynucleotide in a subject.

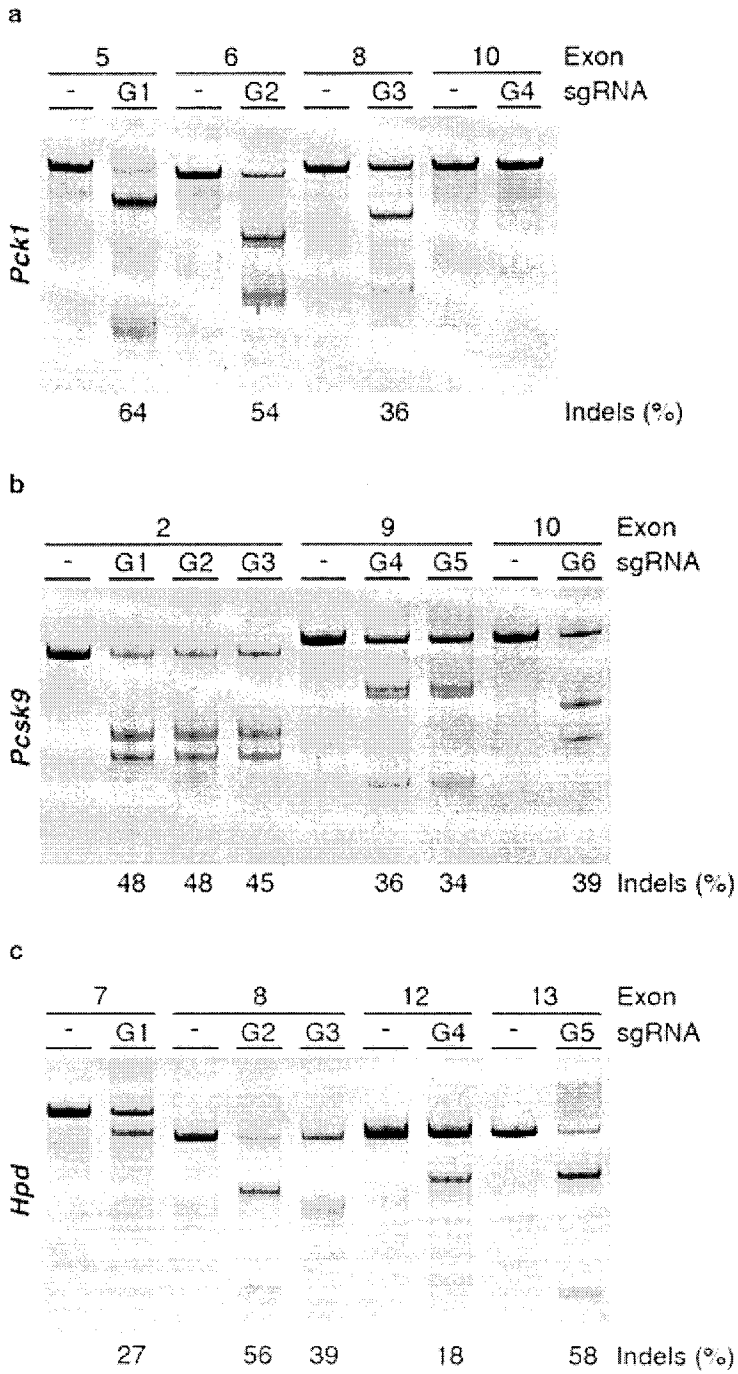
134. The sgRNA, nucleic acid, vector, composition and/or system for use of claim 133, wherein the condition is a metabolic condition.

135. The sgRNA, nucleic acid, vector, composition and/or system for use of claim 133 or 134, wherein the condition is a hepatic condition.

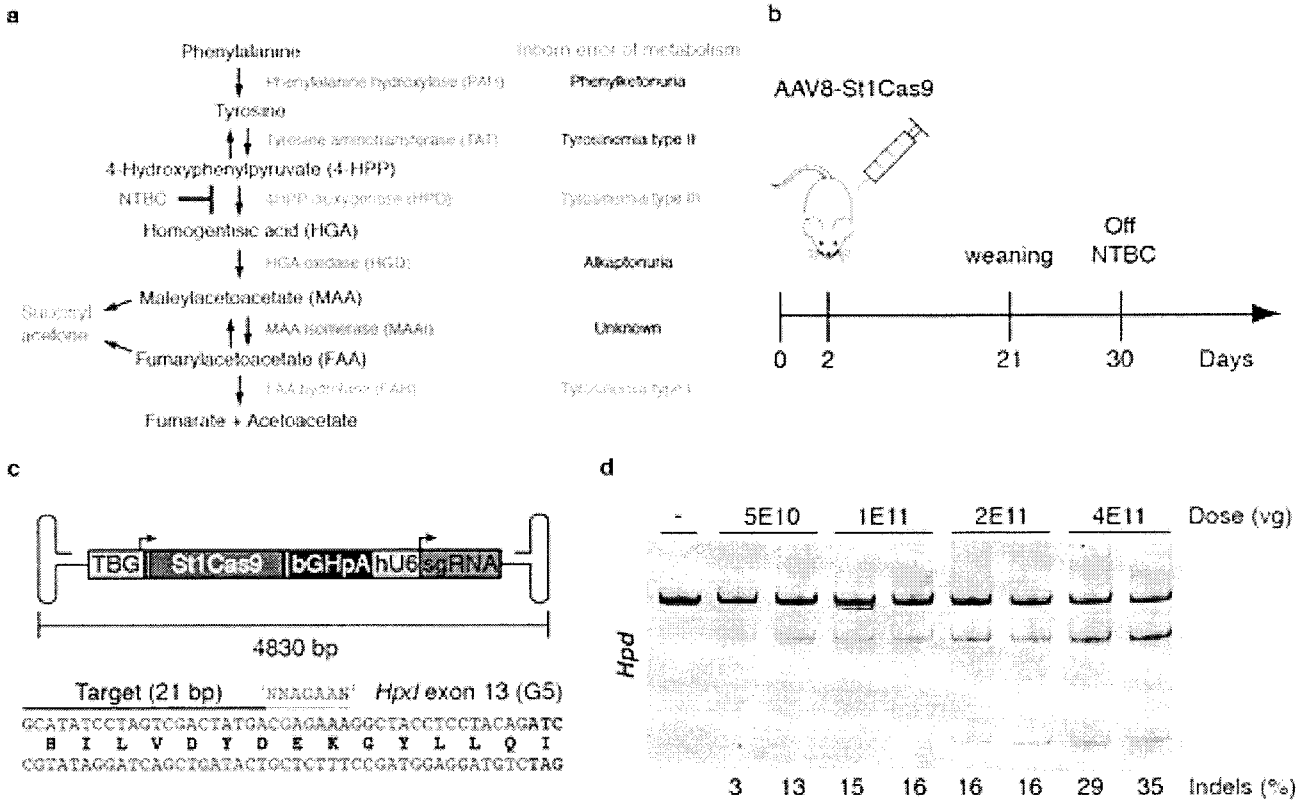
136. The polypeptide according to any one of claims 74 to 103, the nucleic acid of claim 104, the vector of any one of claims 105 to 112, the host cell of claim 115, and/or the composition of claim 116 or 117, and/or the system of claim 118 or 119, for use as a medicament.


a
1121 AA
NLS
St1Cas9 (LMD-9)
NLS
1149 AA
b
Upper stem
Bulge
Stemloop 1
Stemloop 2
Lower stem
Guide
Linker
c
Extended PAM
Target (20 bp)
NNAGAAW
FANCF
GACGTAGGTAGTGCTTGAGACCGCCAGAAGCT
CTGCATCCATCACGAACTCTGGCGGTCTTCGA
EMX1
GAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAA
CTCCTTCCCGGACTCAGGCTCGTCTTCTTCTT
RUNX1
CCAGAGGTATCCAGCAGAGGGGAGAAGAAAGA
GGTCTCCATAGGTCGTCTCCCCTCTTCTTTCT
d
v0
v1
sgRNA
- 50 100 250 50 100 250 ng
FANCF
7 16 31 28 40 55 Indels (%)
4 7 22 21 38 67 TIDE (%)
EMX1
15 20 53 54 56 63 Indels (%)
4 7 32 40 54 85 TIDE (%)
RUNX1
10 14 38 36 36 40 Indels (%)
12 22 53 66 76 90 TIDE (%)

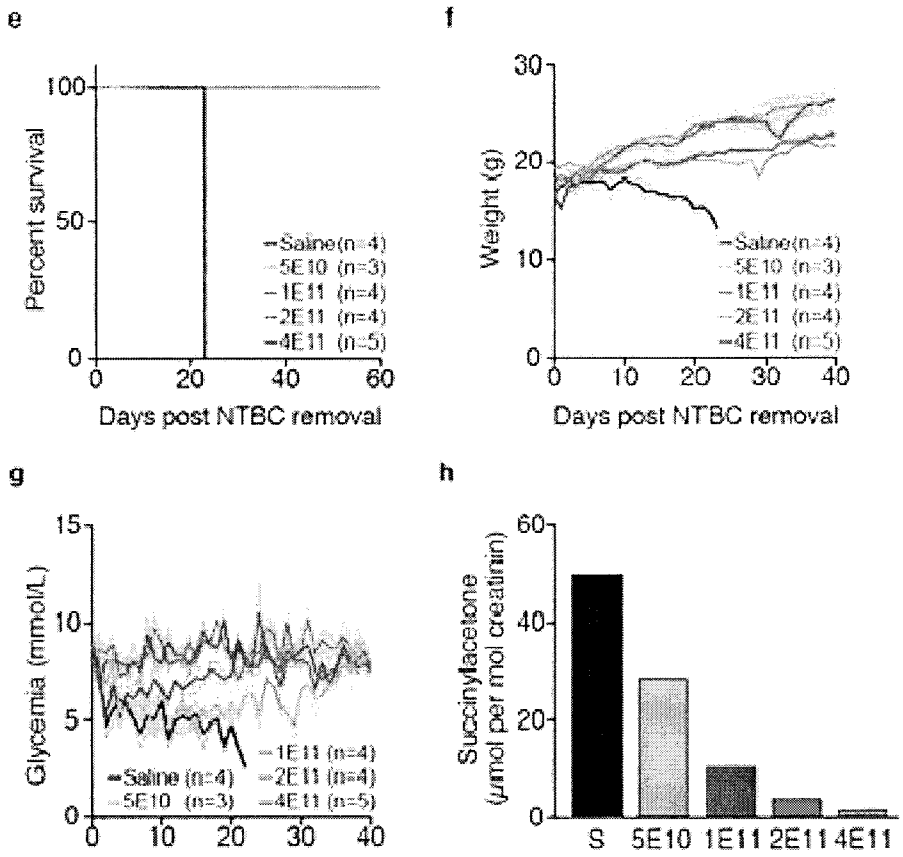
FIGs. 1a-1d


a
5
6
8
10
Exon
-
G1
-
G2
-
G3
-
G4
sgRNA
Pck1
64
54
36
Indels (%)
b
2
9
10
Exon
-
G1
G2
G3
-
G4
G5
-
G6
sgRNA
Pcsk9
48
48
45
36
34
39
Indels (%)
c
7
8
12
13
Exon
-
G1
-
G2
G3
-
G4
-
G5
sgRNA
Hpd
27
56
39
18
58
Indels (%)

FIGs. 2a-2c


a
Phenylalanine
Phenylketonuria
Tyrosine
4-Hydroxyphenylpyruvate (4-HPP)
NTBC
Homogentisic acid (HGA)
Alkaptonuria
Maleylacetoacetate (MAA)
Unknown
Fumarylacetoacetate (FAA)
Fumarate + Acetoacetate
b
AAV8-St1Cas9
weaning
Off NTBC
0 2 21 30 Days
c
TBG St1Cas9 bGHpA hU6 sgRNA
4830 bp
Target (21 bp)
Hpd exon 13 (G5)
GCATATCCTAGTCGACTATGACGAGAAAGGCTACCTCCTACAGATC
H I L V D Y D E K G Y L L Q I
CGTATAGGATCAGCTGATACTGCTCTTTCCGATGGAGGATGTCTAG
d
- 5E10 1E11 2E11 4E11 Dose (vg)
Hpd
3 13 15 16 16 16 29 35 Indels (%)

FIGs. 3a-3d


e
Percent survival
100
50
0
0
20
40
60
Days post NTBC removal
Saline(n=4)
5E10 (n=3)
1E11 (n=4)
2E11 (n=4)
4E11 (n=5)
f
Weight (g)
30
20
10
0
0
10
20
30
40
Days post NTBC removal
Saline(n=4)
5E10 (n=3)
1E11 (n=4)
2E11 (n=4)
4E11 (n=5)
g
Glycemia (mmol/L)
15
10
5
0
0
10
20
30
40
Saline (n=4)
5E10 (n=3)
1E11 (n=4)
2E11 (n=4)
4E11 (n=5)
h
Succinylacetone (µmol per mol creatinin)
60
40
20
0
S
5E10
1E11
2E11
4E11

FIGs. 3e-3h

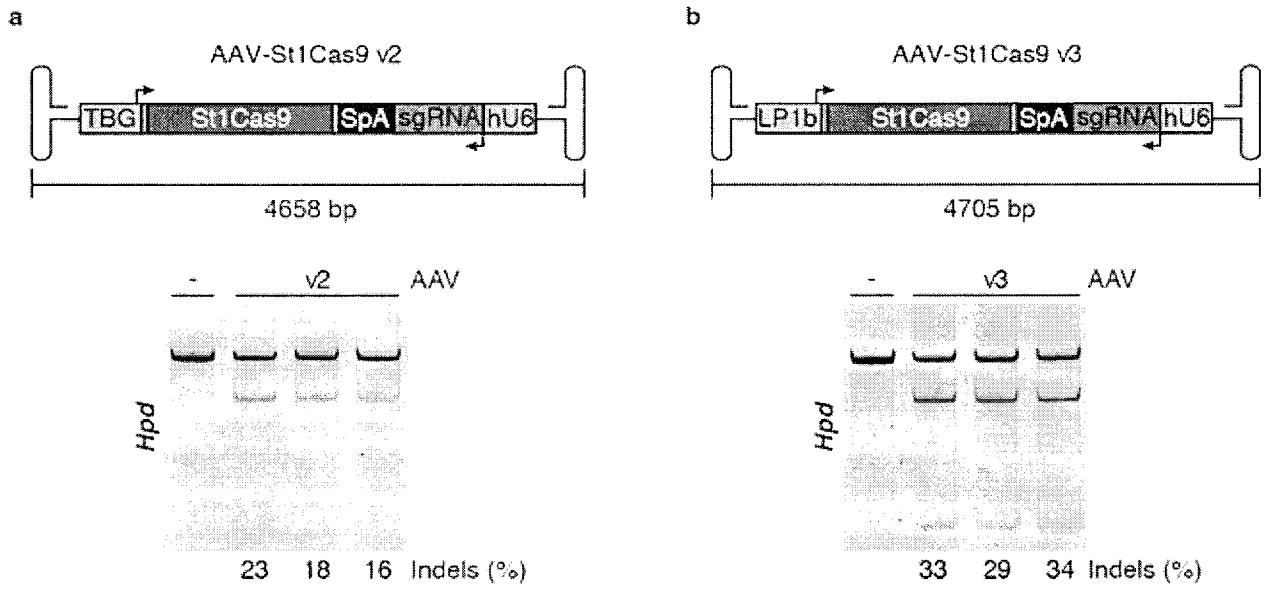
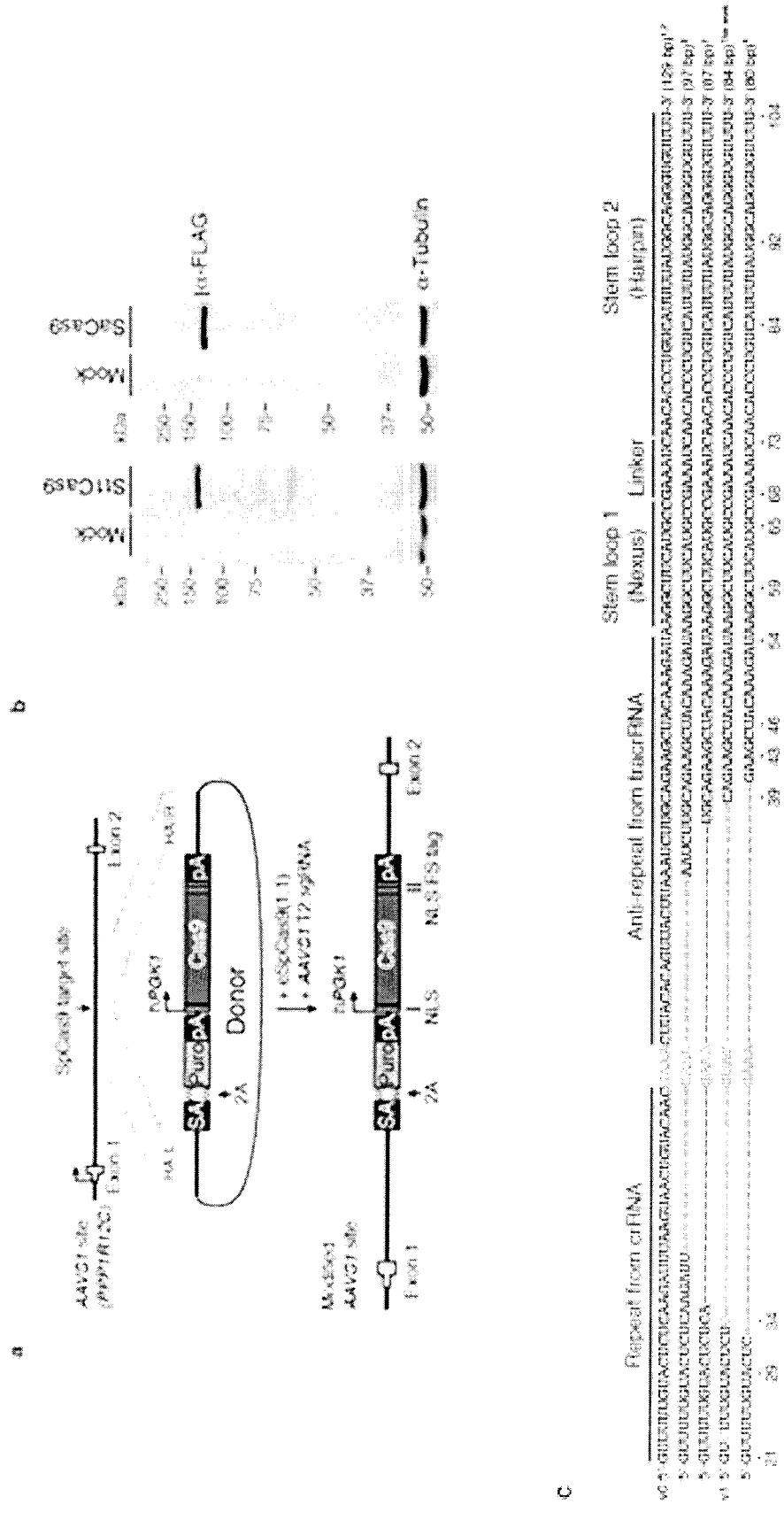

a
AAV-St1Cas9 v2
TBG
St1Cas9
SpA
sgRNA
hU6
4658 bp
-
v2
AAV
Hpd
23 18 16 Indels (%)
b
AAV-St1Cas9 v3
LP1b
St1Cas9
SpA
sgRNA
hU6
4705 bp
-
v3
AAV
Hpd
33 29 34 Indels (%)

FIG. 4


a
AAVS1 site
(PPP1R12C)
Exon 1
SpCas9 target site
Exon 2
SA
PuroR
pA
2A
Donor
hPGK
Cas9
NLS
NLS FS tag
• SpCas9 (1.1)
• AAVS1 T2 sgRNA
Modified
AAVS1 site
b
Mock
SiCas9
SaCas9
kDa
250
150
100
75
50
37
Ki-FLAG
α-Tubulin
c
Repeat from crRNA
Anti-repeat from tracrRNA
Stem loop 1
(Nexus)
Linker
Stem loop 2
(Hairpin)

FIGs. 5a-5c

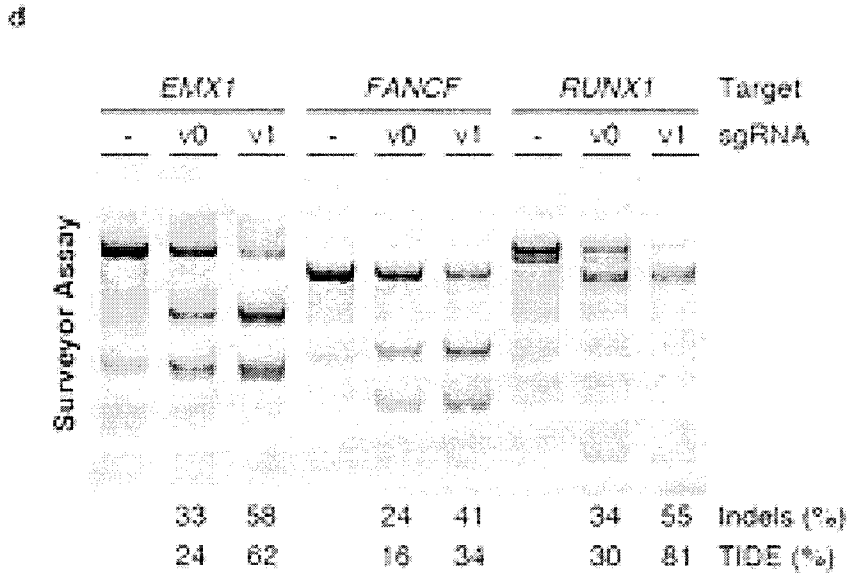
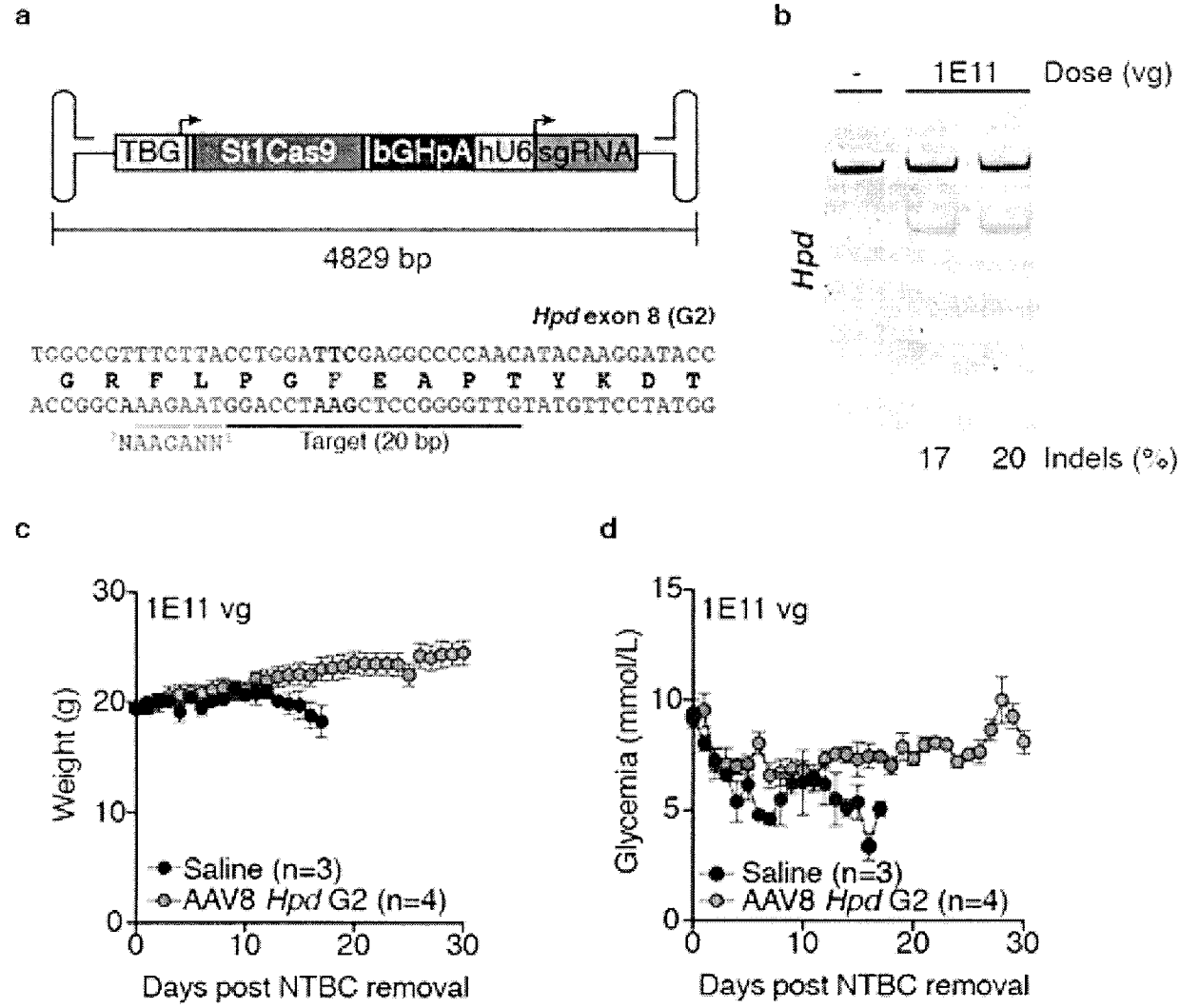

d
EMX1
FANCF
RUNX1
Target
-
v0
v1
-
v0
v1
-
v0
v1
sgRNA
Surveyor Assay
33
58
24
41
34
55
Indels (%)
24
62
16
34
30
81
TIDE (%)

FIG. 5d



FIGs. 6a-6d

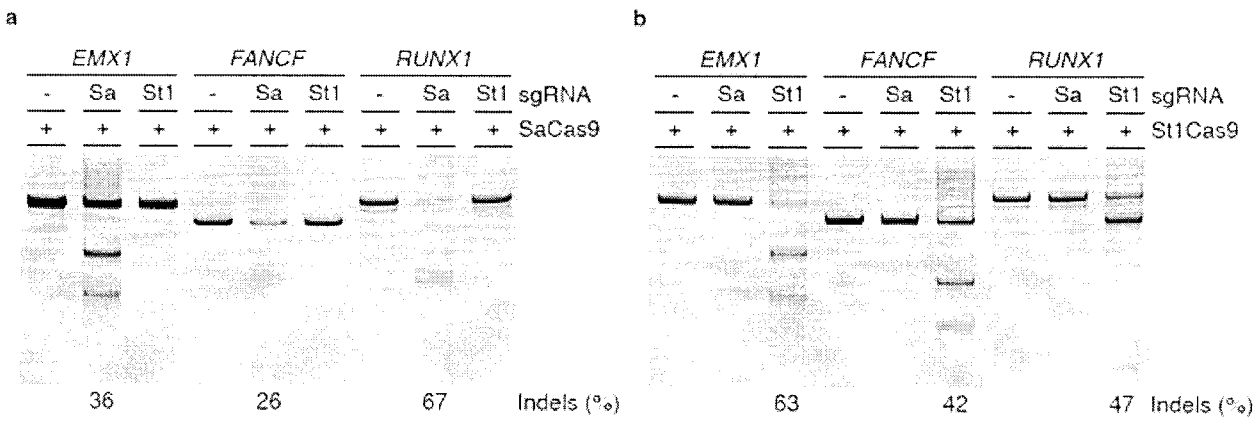


FIG. 7

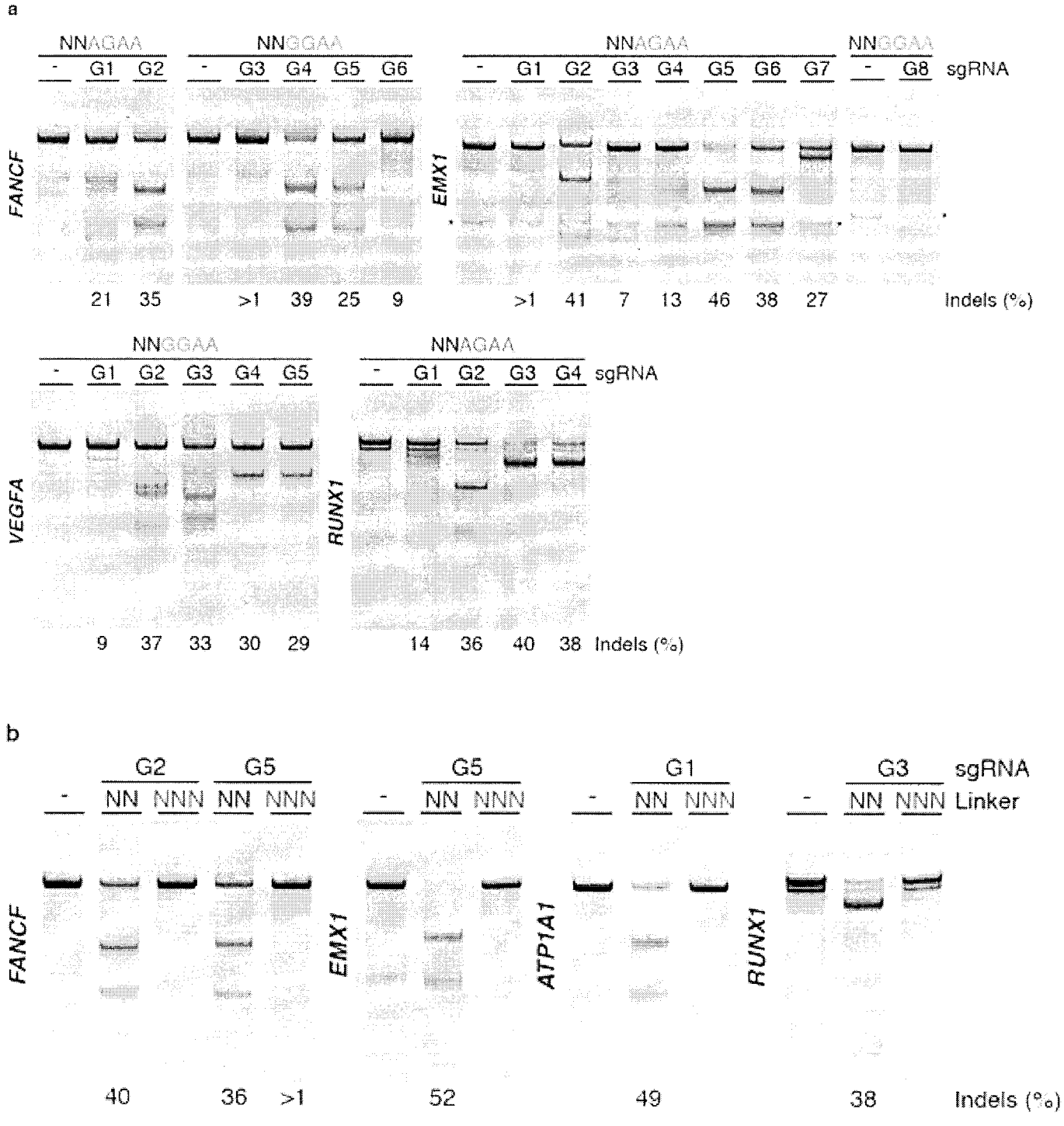

a
NNAGAA
NNGGAA
NNAGAA
NNGGAA
- G1 G2 - G3 G4 G5 G6
- G1 G2 G3 G4 G5 G6 G7 - G8
sgRNA
FANCF
EMX1
21 35 >1 39 25 9
>1 41 7 13 46 38 27
Indels (%)
NNGGAA
NNAGAA
- G1 G2 G3 G4 G5
- G1 G2 G3 G4
sgRNA
VEGFA
RUNX1
9 37 33 30 29
14 36 40 38
Indels (%)
b
G2 G5 G5 G1 G3
- NN NNN NN NNN - NN NNN - NN NNN - NN NNN
sgRNA
Linker
FANCF
EMX1
ATP1A1
RUNX1
40 36 >1 52 49 38
Indels (%)

FIG. 8

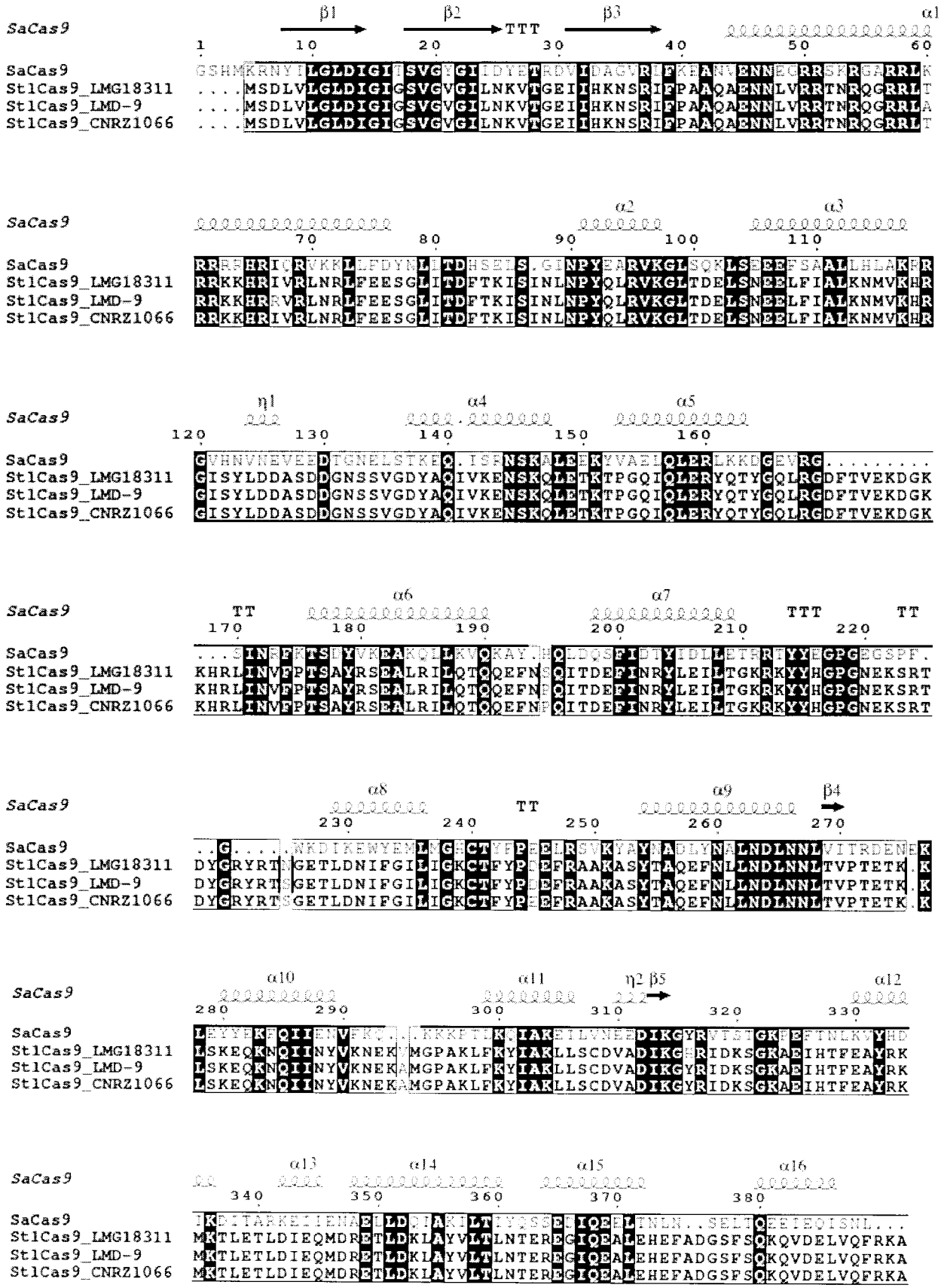


FIG. 9a

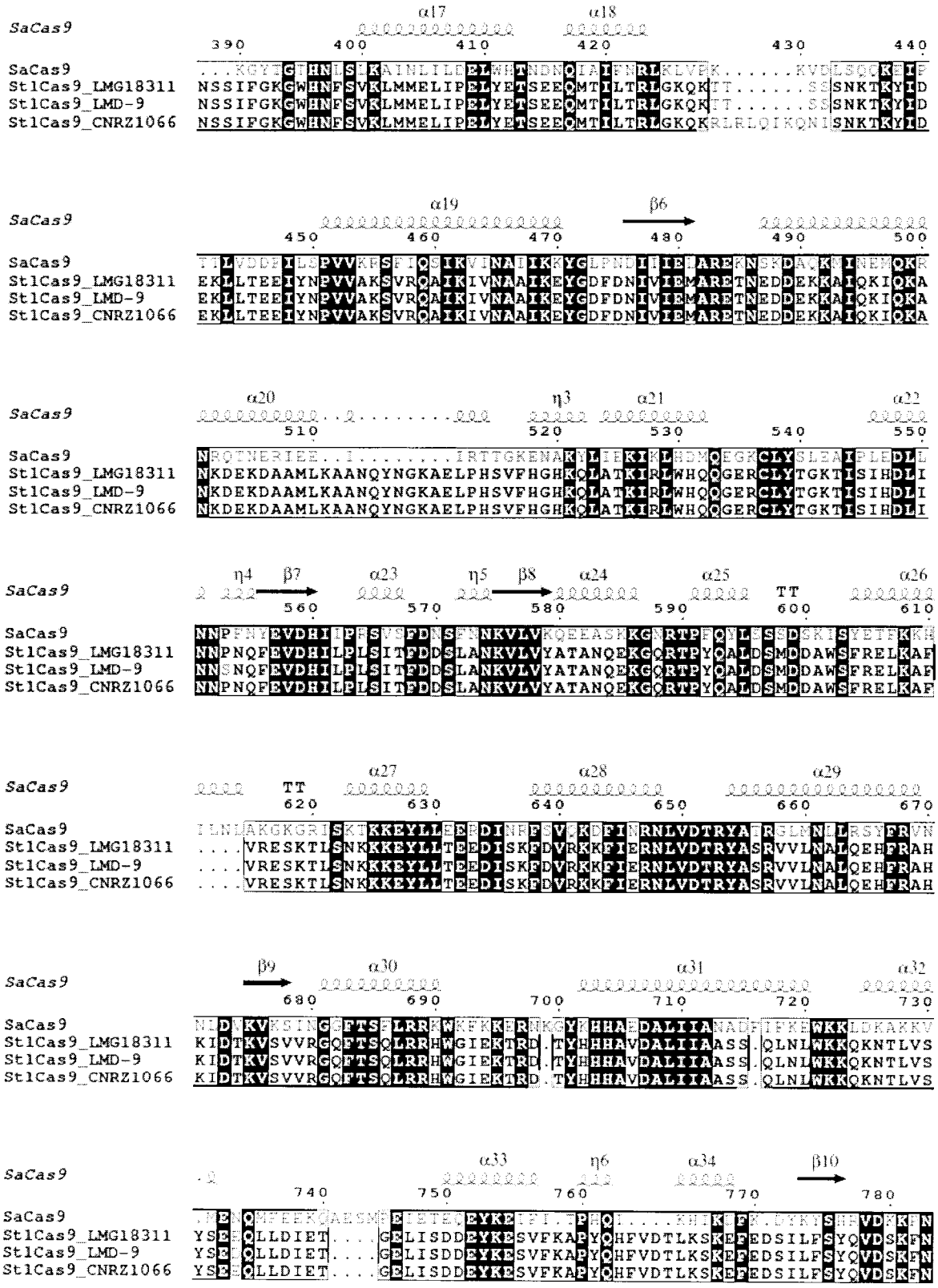


FIG. 9b

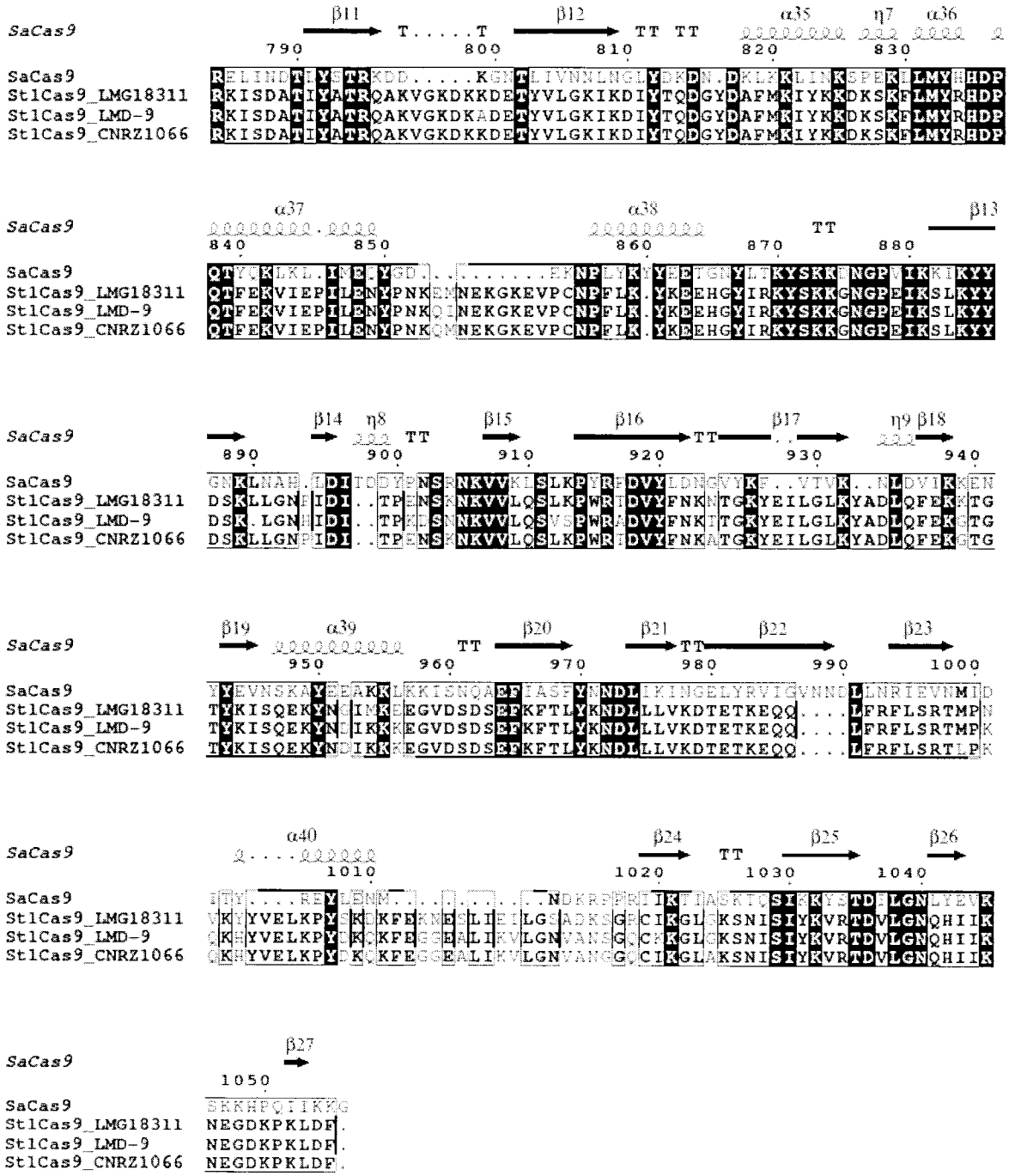

SaCas9
SaCas9
St1Cas9_LMG18311
St1Cas9_LMD-9
St1Cas9_CNRZ1066

FIG. 9c

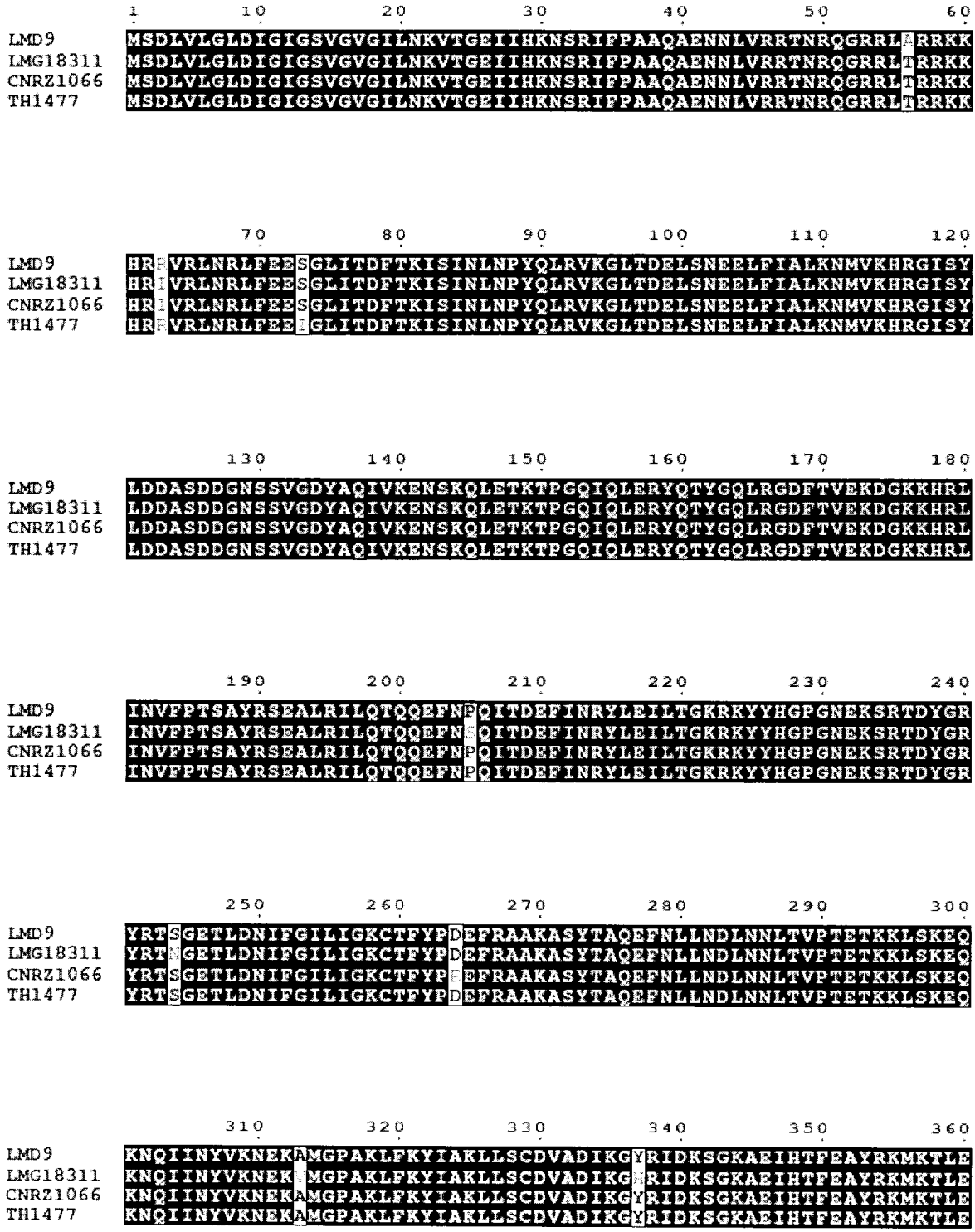

1 10 20 30 40 50 60
LMD9 MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKK
LMG18311 MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLTRRKK
CNRZ1066 MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLTRRKK
TH1477 MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLTRRKK
70 80 90 100 110 120
LMD9 HRPVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISY
LMG18311 HRIVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISY
CNRZ1066 HRIVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISY
TH1477 HRPVRLNRLFEEIGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISY
130 140 150 160 170 180
LMD9 LDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRL
LMG18311 LDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRL
CNRZ1066 LDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRL
TH1477 LDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRL
190 200 210 220 230 240
LMD9 INVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGR
LMG18311 INVFPTSAYRSEALRILQTQQEFNSQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGR
CNRZ1066 INVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGR
TH1477 INVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGR
250 260 270 280 290 300
LMD9 YRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQ
LMG18311 YRTNGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQ
CNRZ1066 YRTSGETLDNIFGILIGKCTFYPEEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQ
TH1477 YRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQ
310 320 330 340 350 360
LMD9 KNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLE
LMG18311 KNQIINYVKNEKTMGPAKLFKYIAKLLSCDVADIKGFRIDKSGKAEIHTFEAYRKMKTLE
CNRZ1066 KNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLE
TH1477 KNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLE

FIG. 10a

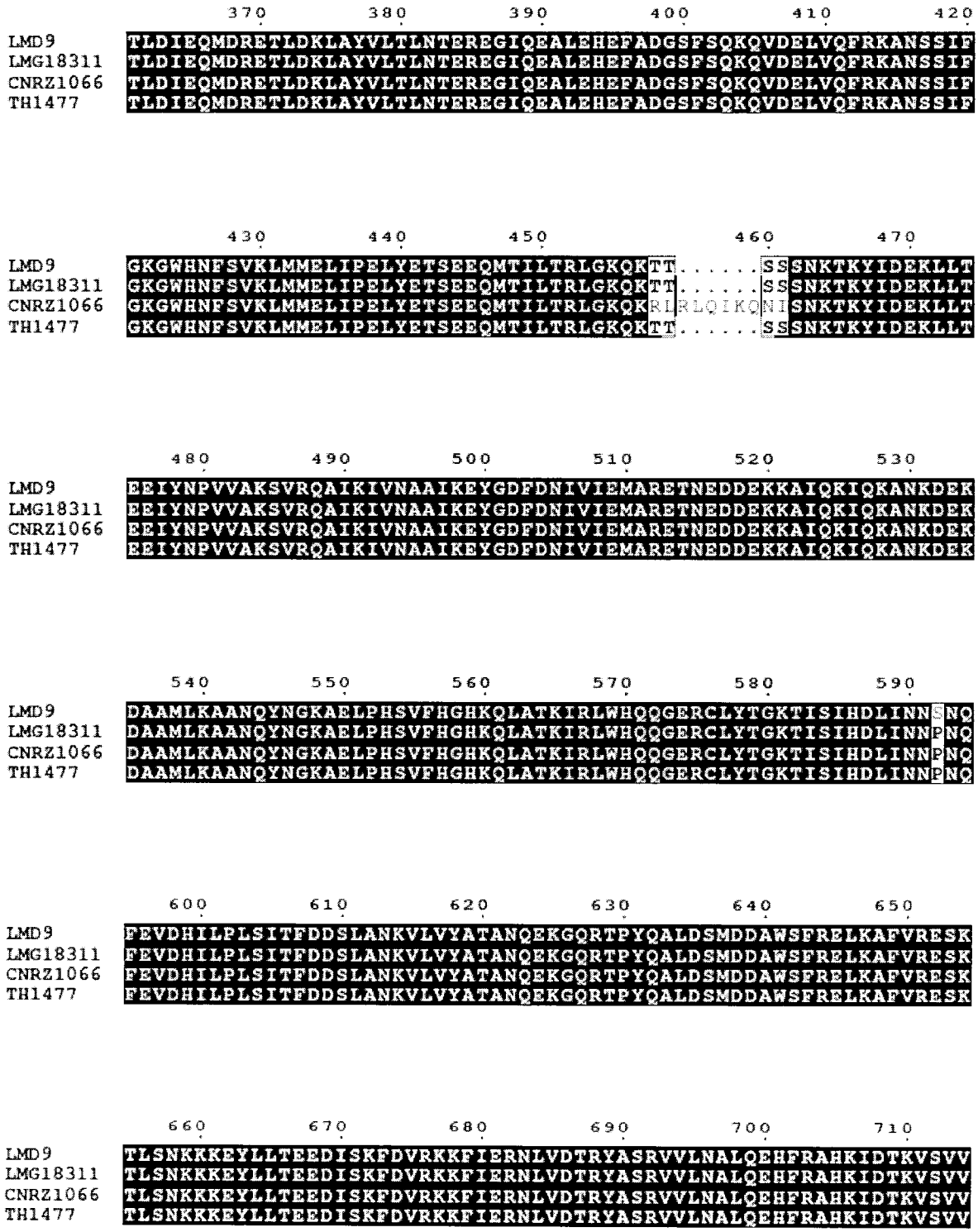

370 380 390 400 410 420
LMD9 TLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIF
LMG18311 TLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIF
CNRZ1066 TLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIF
TH1477 TLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIF
430 440 450 460 470
LMD9 GKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTT.....SSSNKTKYIDEKLLT
LMG18311 GKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTT.....SSSNKTKYIDEKLLT
CNRZ1066 GKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKRLRLQIKQNISNKTKYIDEKLLT
TH1477 GKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTT.....SSSNKTKYIDEKLLT
480 490 500 510 520 530
LMD9 EEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEK
LMG18311 EEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEK
CNRZ1066 EEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEK
TH1477 EEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEK
540 550 560 570 580 590
LMD9 DAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQ
LMG18311 DAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNPNQ
CNRZ1066 DAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNPNQ
TH1477 DAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNPNQ
600 610 620 630 640 650
LMD9 FEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESK
LMG18311 FEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESK
CNRZ1066 FEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESK
TH1477 FEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESK
660 670 680 690 700 710
LMD9 TLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVV
LMG18311 TLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVV
CNRZ1066 TLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVV
TH1477 TLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVV

FIG. 10b

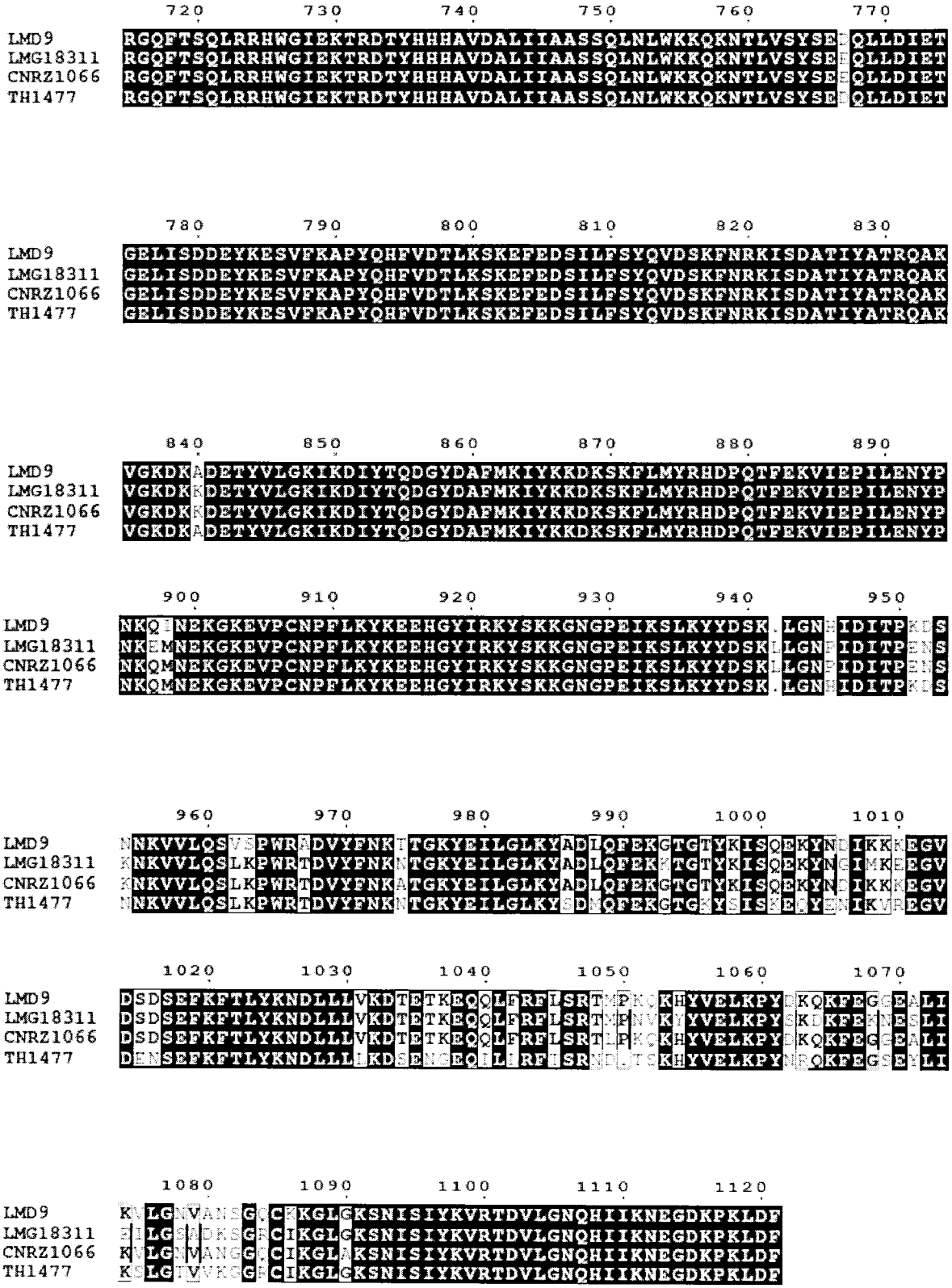
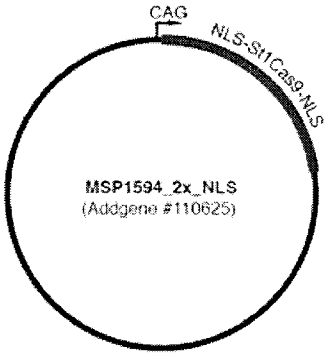

720 730 740 750 760 770
LMD9 RGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSE QLLDIET
LMG18311 RGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSE QLLDIET
CNRZ1066 RGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSE QLLDIET
TH1477 RGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSE QLLDIET
780 790 800 810 820 830
LMD9 GELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAK
LMG18311 GELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAK
CNRZ1066 GELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAK
TH1477 GELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAK
840 850 860 870 880 890
LMD9 VGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYP
LMG18311 VGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYP
CNRZ1066 VGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYP
TH1477 VGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYP
900 910 920 930 940 950
LMD9 NKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSK.LGNHIDITPKES
LMG18311 NKEMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNFIDITPEKS
CNRZ1066 NKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNFIDITPEKS
TH1477 NKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSK.LGNHIDITPKES
960 970 980 990 1000 1010
LMD9 NNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNLIKKKEGV
LMG18311 KNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYADLQFEKKTGTYKISQEKYNGIMKEEGV
CNRZ1066 KNKVVLQSLKPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKISQEKYNIIKKKEGV
TH1477 NNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDVQFEKGTGKYSISKEEYENIKVREGV
1020 1030 1040 1050 1060 1070
LMD9 DSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALI
LMG18311 DSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPNVKYYVELKPYSKDKFEKNESLI
CNRZ1066 DSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQKHYVELKPYDKQKFEGGEALI
TH1477 DENSEFKFTLYKNDLLLIKDSENGEQILIRFISRNMDLTSKHYVELKPYNPQKFEGSEYLI
1080 1090 1100 1110 1120
LMD9 KVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF
LMG18311 EILGSIDKSGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF
CNRZ1066 KVLGYMANGGQCIKGLAKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF
TH1477 KSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF

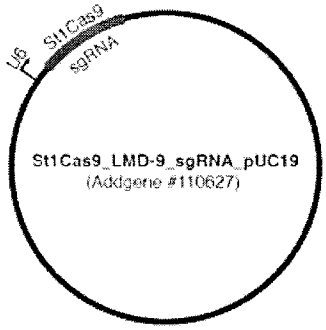
FIG. 10c

Plasmids for *in vitro* genome editing

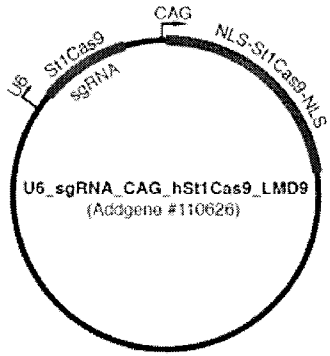
Mammalian expression vector for St1Cas9 LMD-9



hU6-driven sgRNA expression plasmid for St1Cas9 LMD-9

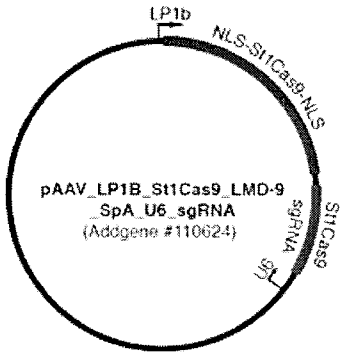


Dual expression vector for St1Cas9 LMD-9 and its sgRNA



Plasmid for *in vivo* genome editing

Single AAV vector with a liver-specific promoter to express St1Cas9 LMD-9 and its hU6-driven sgRNA



Cloning

- The guide $N_{(x)}$ can vary in size between 18 and 22.
- Annealed oligonucleotides must have the indicated overhangs.
- sgRNA expression cassettes are linearized by digestion with BsmBI.

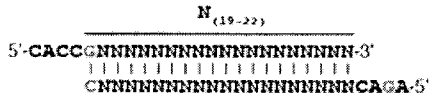


FIG. 11

FANCF				
	Editing efficacy (%)			
sgRNA		Dose (ng)		
	ctrl (-)	50	100	250
v0	0	4	7	22
v1	0	21	38	67

FIG. 12

EMX1				
	Editing efficacy (%)			
sgRNA	Dose (ng)			
	ctrl (-)	50	100	250
v0	0	4	7	32
v1	0	40	54	85

FIG. 13

RUNX1				
	Editing efficacy (%)			
sgRNA	Dose (ng)			
	ctrl (-)	50	100	250
v0	0	12	22	53
v1	0	66	76	90

FIG. 14

Target	sgRNA		
	ctrl (-)	v0	v1
EMX1	0	24	62
FANCF	0	16	34
RUNX1	0	30	81

FIG. 15

St1Cas9 (LMD-9) DNA
SV40 NLS
ST1CAS9 SEQUENCE
linkers

ATGggcgccCCAAAGAAGAAGCGGAAGGTCggtatccacggagtcccagcagccAGCGACCTGGTGCTGGGCCTGGACAT
CGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTGACCGGCGAGATCATCCACAAGAACAGTCGCATCTTCCCTG
CTGCTCAGGCTGAGAACAACCTGGTGCGCCGCACCAACCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGC
GTGCGCCTGAACCGCCTGTTCGAGGAGAGCGGCCTGATCACCGACTTCACCAAGATCAGCATCAACCTGAACCCCTACCA
GCTGCGCGTGAAGGGCCTGACCGACGAGCTGAGCAACGAGGAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCG
GCATCAGCTACCTGGACGACGCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAACAGC
AAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGCCAGCTGCGCGGCGACTTCAC
CGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTGTTCCCCACCAGCGCCTACCGCAGCGAGGCCCTGCGCATCC
TGCAGACCCAGCAGGAGTTCAACCCCCAGATCACCGACGAGTTCATCAACCGCTACCTGGAGATCCTGACCGGCAAGCGC
AAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCACCGACTACGGCCGCTACCGCACCAGCGGCGAGACCCTGGACAA
CATCTTCGGCATCCTGATCGGCAAGTGCACCTTCTACCCCGACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGG
AGTTCAATCTGCTGAACGACCTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAG
ATCATCAACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATCGCCAAGCTGCTGAGCTGCGA
CGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGCAAGGCCGAGATCCACACCTTCGAGGCCTACCGCAAGATGA
AGACCCTGGAGACCCTGGACATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGGCCTACGTGCTGACCCTGAACACC
GAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGGCAGCTTCAGCCAGAAACAGGTGGACGAGCTGGT
GCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAGGGCTGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCC
CCGAGCTGTACGAGACCAGCGAGGAGCAGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAAC
AAGACCAAGTACATCGACGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCCAAGAGCGTGCGCCAGGCCAT
CAAGATCGTGAACGCCGCCATCAAGGAGTACGGCGACTTCGACAACATCGTGATCGAGATGGCCCGCGAGACCAACGAGG
ACGACGAGAAGAAGGCCATCCAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCTGAAGGCCGCCAAC
CAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTGGCCACCAAGATCCGCCTGTGGCA
CCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATCAGCATCCACGACCTGATCAACAACAGCAACCAGTTCGAGG
TGGACCACATCCTGCCCCTGAGCATCACCTTCGACGACAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAG
GAGAAGGGCCAGCGCACCCCCTACCAGGCCCTGGACAGCATGGACGACGCCTGGAGCTTCCGCGAGCTGAAGGCCTTCGT
GCGCGAGAGCAAGACCCTGAGCAACAAGAAGAAGGAGTATCTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCA
AGAAGTTCATCGAGCGCAACCTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTCCGC
GCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGCCGCCACTGGGGCATCGAGAA
GACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTGATCATTGCGGCTTCTAGCCAGCTGAACCTGTGGAAGAAGC
AGAAGAACACCCTGGTGAGCTACAGCGAGGACCAGCTGCTGGACATCGAGACCGGCGAGCTGATCAGCGACGACGAGTAC
AAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCGTGGACACCCTGAAGAGCAAGGAGTTCGAGGACAGCATCCTGTT
CAGCTACCAGGTGGACAGCAAGTTCAACCGCAAGATCAGCGACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCA
AGGACAAGGCCGACGAGACCTACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAG
ATCTACAAGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAGAAGGTGATCGAGCCCATCCT
GGAGAACTACCCCAACAAGCAGATCAACGAGAAAGGCAAGGAGGTGCCCTGCAACCCCTTCCTGAAGTACAAGGAGGAGC
ACGGCTACATCCGCAAGTACAGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGAAGTACTACGACAGCAAGCTGGGC
AACCACATCGACATCACCCCCAAGGACAGCAACAACAAGGTGGTGCTGCAGAGCGTGAGCCCCTGGCGCGCCGACGTGTA
CTTCAACAAGACCACCGGCAAGTACGAGATCCTGGGGCTGAAGTACGCCGATCTGCAGTTTGAGAAAGGCACAGGCACCT
ACAAGATCAGCCAGGAGAAGTACAACGACATCAAGAAGAAGGACGGCGTGGACAGCGACAGCGAGTTCAAGTTCACCCTG
TACAAGAACGACCTTCTGCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTCCGCTTCCTGAGCCGCACCATGCC
CAAGCAGAAGCACTACGTGGAGCTGAAGCCCTACGACAAGCAGAAGTTCGAGGGCGGCGAGGCCCTGATCAAGGTGCTGG
GCAACGTGGCCAACAGCGGCCAGTGCAAGAAGGGCCTGGGCAAGAGCAACATCAGCATCTACAAGGTGCGCACCGACGTG
CTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTCagcagggctgacCCCAAGAAGAAGAG
GAAGGTGTGA

FIG. 16

St1Cas9 (LMD-9) protein

SV40 NLS

ST1CAS9 SEQUENCE

linkers

MgaPKKKRKVgihgvpaa*SDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRR*
VRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENS
KQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKR
KYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQ
IINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNT
EREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSN
KTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAAN
QYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQ
EKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFR
AHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEY
KESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMK
IYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLG
NHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTL
YKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDV
LGNQHIIKNEGDKPKLDFsradPKKKRKV

FIG. 17

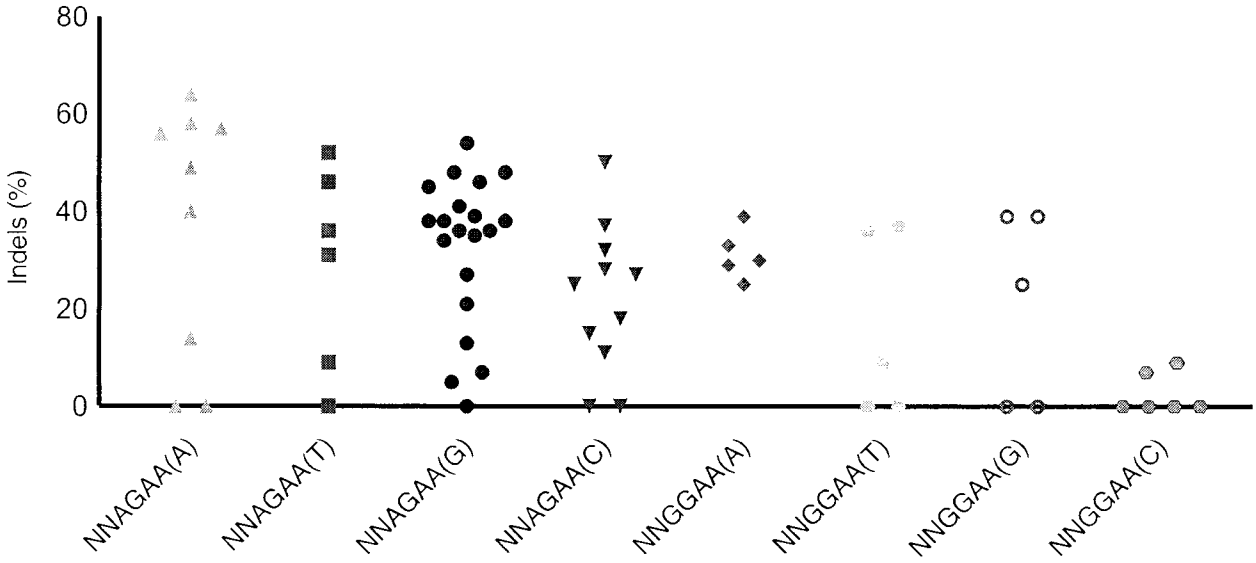


FIG. 18

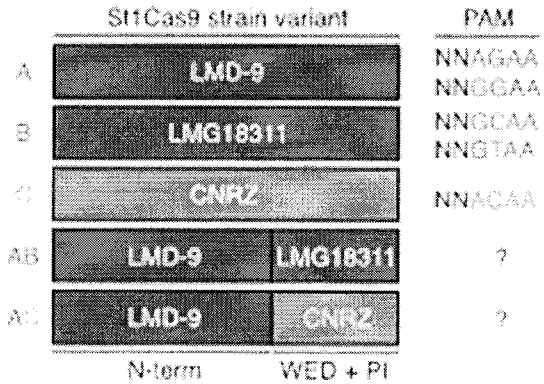
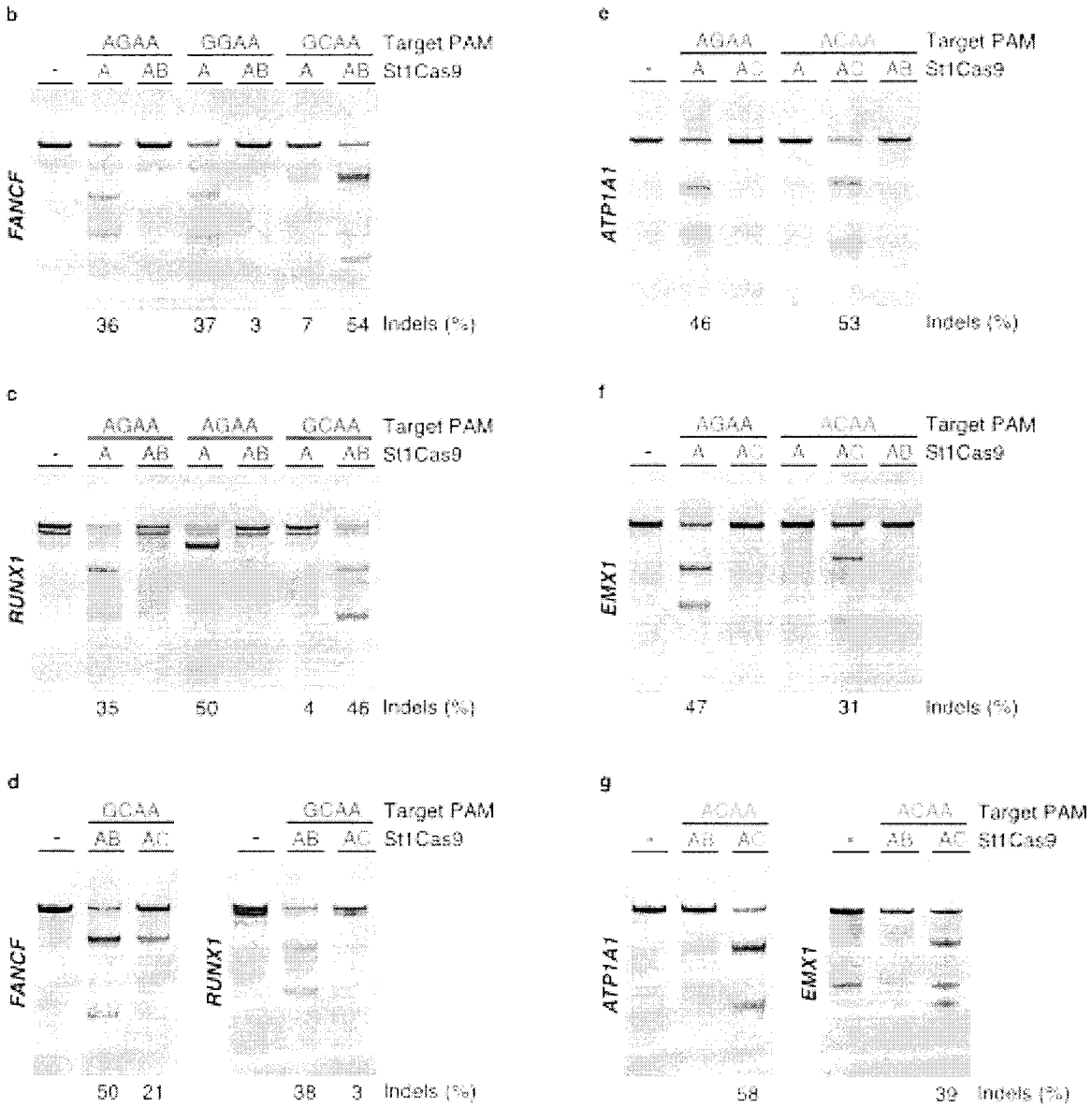
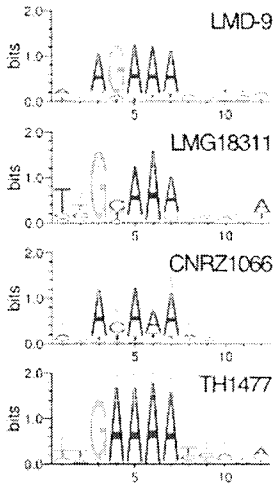


FIG. 19a

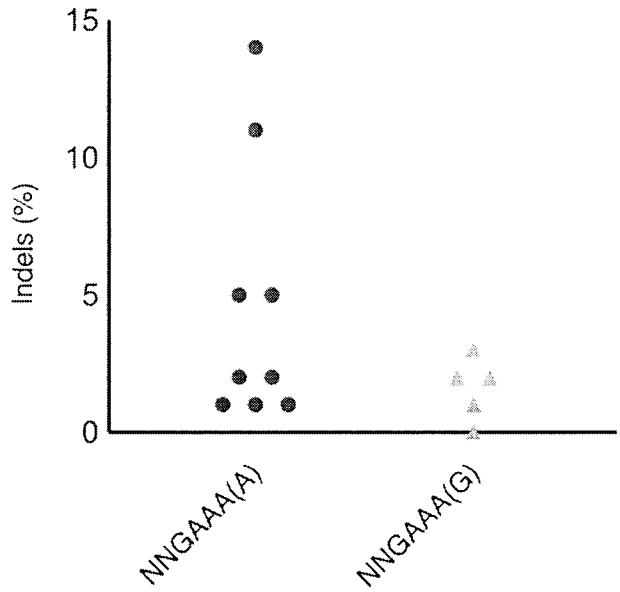

b
AGAA
GGAA
GCAA
Target PAM
-
A
AB
A
AB
A
AB
St1Cas9
FANCF
36
37
3
7
54
Indels (%)
c
AGAA
AGAA
GCAA
Target PAM
St1Cas9
RUNX1
35
50
4
46
Indels (%)
d
GCAA
GCAA
Target PAM
AB
AC
St1Cas9
FANCF
RUNX1
50
21
38
3
Indels (%)
e
AGAA
ACAA
Target PAM
A
AC
AB
St1Cas9
ATP1A1
46
53
Indels (%)
f
AGAA
ACAA
Target PAM
St1Cas9
EMX1
47
31
Indels (%)
g
ACAA
ACAA
Target PAM
St1Cas9
ATP1A1
EMX1
58
39
Indels (%)

FIGs. 19b-19g

a



b



FIGs. 20a-20b

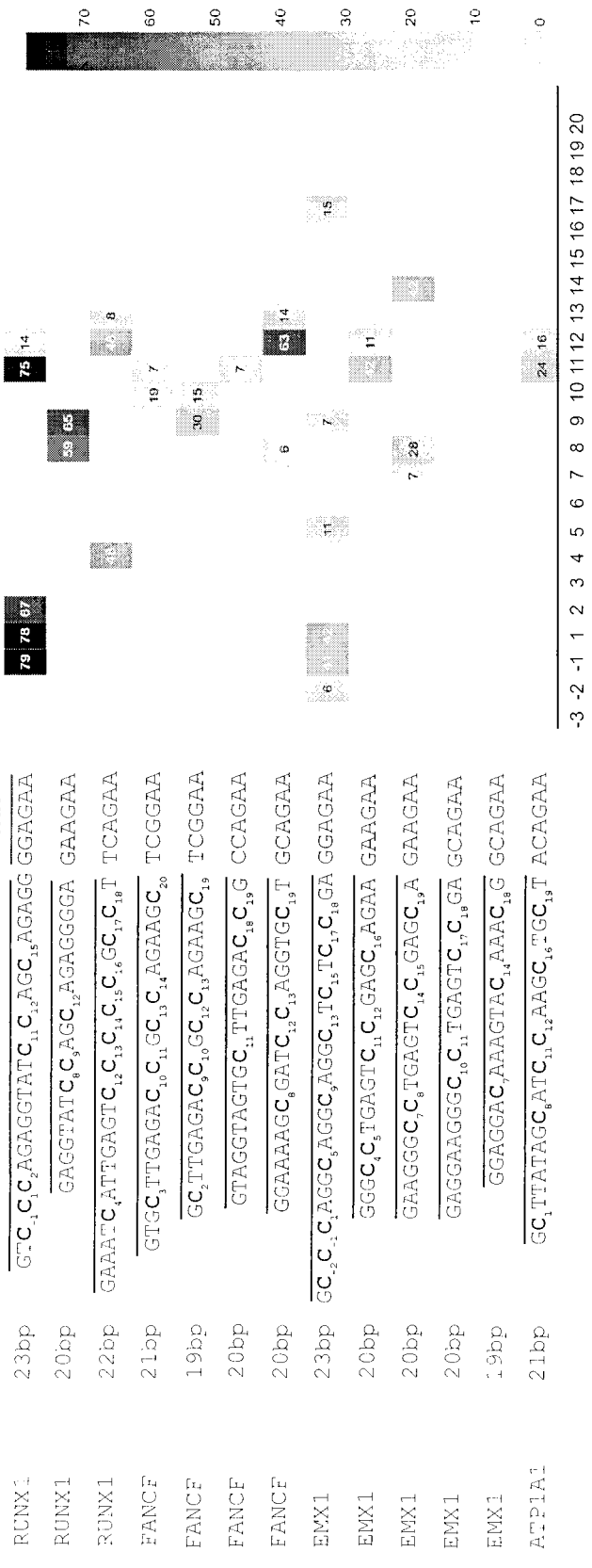
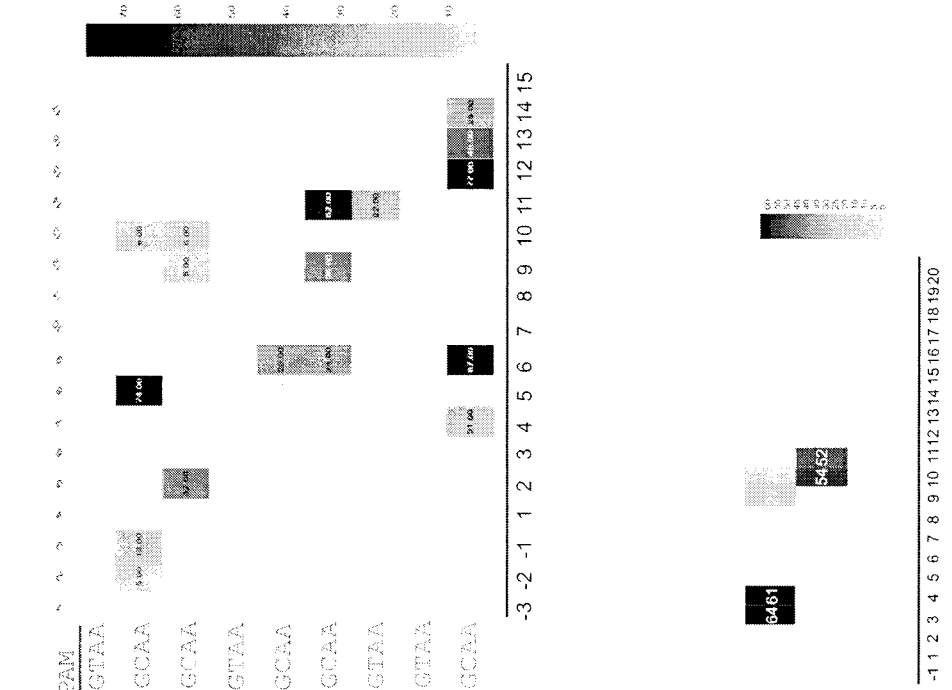

Protospacer PAM
RUNX1 23bp GTCC-1C1C2AGAGGTATC11C12AGC15AGAGG GGAGAA
RUNX1 20bp GAGGTATC8C9AGC12AGAGGGGA GAAGAA
RUNX1 22bp GAAATTC4ATTGAGTC12C13C14C15C16GC17C18T TCAGAA
FANCF 21bp GTGC3TTGAGAC10C11GC13C14AGAAGC20 TCGGAA
FANCF 19bp GC2TTGAGAC9C10GC12C13AGAAGC19 TCGGAA
FANCF 20bp GTAGGTAGTGC11TTGAGAC18C19G CCAGAA
FANCF 20bp GGAAAAGC8GATC12C13AGGTGC19T GCAGAA
EMX1 23bp GC-2C-1C1AGGC5AGGC9AGGC13TC15TC17C18GA GGAGAA
EMX1 20bp GGGC4C5TGAGTC11C12GAGC16AGAA GAAGAA
EMX1 20bp GAAGGGC7C8TGAGTC14C15GAGC19A GAAGAA
EMX1 20bp GAGGAAGGGC10C11TGAGTC17C18GA GCAGAA
EMX1 19bp GGAGGAC7AAAGTAC14AAAC18G GCAGAA
ATP1A1 21bp GC1TTATAGC8ATC11C12AAGC16TGC19T ACAGAA
79 75 67
75 14
59 65
8
19 7
30 15
7
6
63 14
6 11 7 15
11
7 28
24 16
70 60 50 40 30 20 10 0
-3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

FIG. 21


a
Protospacer
PAM
RUNX1 20bp GAGGTGAGTAC11ATGC15TGGTC20 TTGTAA
RUNX1 23bp GC-2C-1ATTTC5ATTAC10AGGC14AAAGC19T GACCAA
RUNX1 23bp GAAAC2AAGC6TGC9C10ATTTC25ATTAC20 AGGCAA
Grin2B 20bp GTTAAAATAGGATC14TAC17ATC20 ACGTAA
Grin2B 20bp GGC3TTC6C7TGGTC12TGTGTC16AT CAGCAA
FANCF 20bp GGC3GGC6TGC9AC11AAC14C15AGTGG AGGCAA
ATP1A1 20bp GTAATGAGTTC11ATAAC16C17ATT AAGTAA
ATP1A1 23bp GTTC1ATAAC5C7ATTAAGTAATGAG TGGTAA
AAVS1 20bp GGAC4AC6AGGATC12C13C14TGGAGG CAGCAA
-3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
b
Protospacer
PAM
EMX1 21bp GTTC3C4AGAAC9C10GGAGGAC17AAA GTACAA
EMX1 20bp GGTGTGGTTC10C11AGAAC16C17GGA GGACAA
ATP1A1 21bp GGAGAAGATATC11TGATGTGTA CTACAA
6461
54.52
-1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

FIGs. 22a-22b

NLS-St1Cas9 LMD-9/LMG18311 Hybrid-NLS ORF DNA sequence
SV40 NLS
ST1CAS9 SEQUENCE
LINKERS

ATGGGCGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC*AGCGAC*
CTGGTGCTGGGCCTGGACATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTG
ACCGGCGAGATCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAAC
CTGGTGCGCCGCACCAACCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGC
GTGCGCCTGAACCGCCTGTTCGAGGAGAGCGGCCTGATCACCGACTTCACCAAGATCAGC
ATCAACCTGAACCCCTACCAGCTGCGCGTGAAGGGCCTGACCGACGAGCTGAGCAACGAG
GAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCGGCATCAGCTACCTGGACGAC
GCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAACAGC
AAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGC
CAGCTGCGCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTG
TTCCCCACCAGCGCCTACCGCAGCGAGGCCCTGCGCATCCTGCAGACCCAGCAGGAGTTC
AACCCCCAGATCACCGACGAGTTCATCAACCGCTACCTGGAGATCCTGACCGGCAAGCGC
AAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCACCGACTACGGCCGCTACCGCACC
AGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTGCACCTTCTACCCC
GACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAACGAC
CTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAG
ATCATCAACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATC
GCCAAGCTGCTGAGCTGCGACGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGC
AAGGCCGAGATCCACACCTTCGAGGCCTACCGCAAGATGAAGACCCTGGAGACCCTGGAC
ATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGGCCTACGTGCTGACCCTGAACACC
GAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGGCAGCTTCAGCCAG
AAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAGGGC
TGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAGCTGTACGAGACCAGC
GAGGAGCAGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAAC
AAGACCAAGTACATCGACGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCC
AAGAGCGTGCGCCAGGCCATCAAGATCGTGAACGCCGCCATCAAGGAGTACGGCGACTTC
GACAACATCGTGATCGAGATGGCCCGCGAGACCAACGAGGACGACGAGAAGAAGGCCATC
CAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCTGAAGGCCGCCAAC
CAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTGGCC
ACCAAGATCCGCCTGTGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATC
AGCATCCACGACCTGATCAACAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTG
AGCATCACCTTCGACGACAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAG
GAGAAGGGCCAGCGCACCCCCTACCAGGCCCTGGACAGCATGGACGACGCCTGGAGCTTC
CGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGACCCTGAGCAACAAGAAGAAGGAGTAT
CTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCAAGAAGTTCATCGAGCGCAAC

FIG. 23a

CTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTCCGC
GCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGC
CGCCACTGGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTG
ATCATTGCGGCTTCTAGCCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGC
TACAGCGAGGAGCAGCTGCTGGACATCGAGACCGGCGAGCTGATCAGCGACGACGAGTAC
AAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCGTGGACACCCTGAAGAGCAAGGAG
TTCGAGGACAGCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAACCGCAAGATCAGC
GACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGAAGGACGAGACC
TACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAG
ATCTACAAGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAG
AAGGTGATCGAGCCCATCCTGGAGAACTACCCCAACAAGCAGATGAACGAGAAAGGCAAG
GAGGTGCCCTGCAACCCCTTCCTGAAGTACAAGGAGGAGCACGGCTACATCCGCAAGTAC
AGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGAAGTACTACGACAGCAAGCTGCTG
GGCAACCCCATCGACATCACCCCCGAAAACAGCAAGAACAAGGTGGTGCTGCAGAGCCTT
AAGCCCTGGCGCACCGACGTGTACTTCAACAAGAACACCGGCAAGTACGAGATCCTGGGG
CTGAAGTACGCCGATCTGCAGTTTGAGAAAAAGACAGGCACCTACAAGATCAGCCAGGAG
AAGTACAACGGCATCATGAAGGAGGAGGGCGTGGACAGCGACAGCGAGTTCAAGTTCACC
CTGTACAAGAACGACCTTCTGCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTC
CGCTTCCTGAGCCGCACCATGCCCAACGTGAAGTACTACGTGGAGCTGAAGCCCTACAGC
AAGGACAAGTTCGAGAAGAACGAGAGCCTGATCGAGATCCTGGGCAGCGCCGACAAGAGC
GGCAGGTGCATCAAGGGCCTGGGCAAGAGCAACATCAGCATCTACAAGGTGCGCACCGAC
GTGCTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTCAGC
AGGGCTGACCCCAAGAAGAAGAGGAAGGTGGGATCC

FIG 23b

NLS-St1Cas9 LMD-9/LMG18311 Hybrid-NLS AA sequence
SV40 NLS
ST1CAS9 SEQUENCE
LINKERS

MGAPKKKRKVGIHGVPAA*SDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGR*
RLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDD
ASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEA
LRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYP
DEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVAD
IKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQ
KQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLL
TEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAAN
QYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLAN
KVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERN
LVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNL
WKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKIS
DATIYATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENY
PNKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSL
KPWRTDVYFNKNTGKYEILGLKYADLQFEKKTGTYKISQEKYNGIMKEEGVDSDSEFKFTLYKNDLLLVK
DTETKEQQLFRFLSRTMPNVKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRTD
*VLGNQHIIKNEGDKPKLDF*SRADPKKKRKVGS

FIG. 24

NLS-St1Cas9 LMD-9/CNRZ1066 Hybrid-NLS ORF DNA sequence
SV40 NLS
ST1CAS9 SEQUENCE
LINKERS

ATGGGCGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAGCGAC
CTGGTGCTGGGCCTGGACATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTG
ACCGGCGAGATCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAAC
CTGGTGCGCCGCACCAACCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGC
GTGCGCCTGAACCGCCTGTTCGAGGAGAGCGGCCTGATCACCGACTTCACCAAGATCAGC
ATCAACCTGAACCCCTACCAGCTGCGCGTGAAGGGCCTGACCGACGAGCTGAGCAACGAG
GAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCGGCATCAGCTACCTGGACGAC
GCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAACAGC
AAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGC
CAGCTGCGCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTG
TTCCCCACCAGCGCCTACCGCAGCGAGGCCCTGCGCATCCTGCAGACCCAGCAGGAGTTC
AACCCCCAGATCACCGACGAGTTCATCAACCGCTACCTGGAGATCCTGACCGGCAAGCGC
AAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCACCGACTACGGCCGCTACCGCACC
AGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTGCACCTTCTACCCC
GACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAACGAC
CTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAG
ATCATCAACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATC
GCCAAGCTGCTGAGCTGCGACGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGC
AAGGCCGAGATCCACACCTTCGAGGCCTACCGCAAGATGAAGACCCTGGAGACCCTGGAC
ATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGGCCTACGTGCTGACCCTGAACACC
GAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGGCAGCTTCAGCCAG
AAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAGGGC
TGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAGCTGTACGAGACCAGC
GAGGAGCAGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAAC
AAGACCAAGTACATCGACGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCC
AAGAGCGTGCGCCAGGCCATCAAGATCGTGAACGCCGCCATCAAGGAGTACGGCGACTTC
GACAACATCGTGATCGAGATGGCCCGCGAGACCAACGAGGACGACGAGAAGAAGGCCATC
CAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCTGAAGGCCGCCAAC
CAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTGGCC
ACCAAGATCCGCCTGTGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATC
AGCATCCACGACCTGATCAACAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTG
AGCATCACCTTCGACGACAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAG
GAGAAGGGCCAGCGCACCCCCTACCAGGCCCTGGACAGCATGGACGACGCCTGGAGCTTC
CGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGACCCTGAGCAACAAGAAGAAGGAGTAT
CTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCAAGAAGTTCATCGAGCGCAAC
CTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTCCGC
GCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGC
CGCCACTGGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTG
ATCATTGCGGCTTCTAGCCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGC
TACAGCGAGGAGCAGCTGCTGGACATCGAGACCGGCGAGCTGATCAGCGACGACGAGTAC

FIG. 25a

AAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCGTGGACACCCTGAAGAGCAAGGAG
TTCGAGGACAGCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAACCGCAAGATCAGC
GACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGAAGGACGAGACC
TACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAG
ATCTACAAGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAG
AAGGTGATCGAGCCCATCCTGGAGAACTACCCCAACAAGCAGATGAACGAGAAAGGCAAG
GAGGTGCCCTGCAACCCCTTCCTGAAGTACAAGGAGGAGCACGGCTACATCCGCAAGTAC
AGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGAAGTACTACGACAGCAAGCTGCTG
GGCAACCCCATCGACATCACCCCCGAAAACAGCAAGAACAAGGTGGTGCTGCAGAGCCTT
AAGCCCTGGCGCACCGACGTGTACTTCAACAAGGCCACCGGCAAGTACGAGATCCTGGGG
CTGAAGTACGCCGATCTGCAGTTTGAGAAAGGCACAGGCACCTACAAGATCAGCCAGGAG
AAGTACAACGACATCAAGAAGAAGGAGGGCGTGGACAGCGACAGCGAGTTCAAGTTCACC
CTGTACAAGAACGACCTTCTGCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTC
CGCTTCCTGAGCCGCACCCTGCCCAAGCAGAAGCACTACGTGGAGCTGAAGCCCTACGAC
AAGCAGAAGTTCGAGGGCGGCGAGGCCCTGATCAAGGTGCTGGGCAACGTGGCCAACGGC
GGCCAGTGCATCAAGGGCCTGGCCAAGAGCAACATCAGCATCTACAAGGTGCGCACCGAC
GTGCTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTCAGC
AGGGCTGACCCCAAGAAGAAGAGGAAGGTGGGATCC

FIG. 25b

NLS-St1Cas9 LMD-9/CNRZ1066 Hybrid-NLS AA sequence
SV40 NLS
ST1CAS9 SEQUENCE
LINKERS

MGAPKKKRKVGIHGVPAA*SDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGR*
RLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDD
ASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEA
LRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYP
DEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVAD
IKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQ
KQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLL
TEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAAN
QYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLAN
KVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERN
LVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNL
WKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKIS
DATIYATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENY
PNKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSL
KPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVK
DTETKEQQLFRFLSRTLPKQKHYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTD
*VLGNQHIIKNEGDKPKLDF*SRADPKKKRKVGS

FIG. 26

NLS-St1Cas9 LMD-9/TH1477 Hybrid-NLS ORF DNA sequence

SV40 NLS

ST1CAS9 SEQUENCE

LINKERS

ATGGGCGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC*AGCGACCTGGT*
GCTGGGCCTGGACATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTGACCGGCGAGA
TCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAACCTGGTGCGCCGCACC
AACCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGCGTGCGCCTGAACCGCCTGTT
CGAGGAGAGCGGCCTGATCACCGACTTCACCAAGATCAGCATCAACCTGAACCCCTACCAGCTGC
GCGTGAAGGGCCTGACCGACGAGCTGAGCAACGAGGAGCTGTTCATCGCCCTGAAGAACATGGTG
AAGCACCGCGGCATCAGCTACCTGGACGACGCCAGCGACGACGGCAACAGCAGCGTGGGCGACTA
CGCCCAGATCGTGAAGGAGAACAGCAAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGG
AGCGCTACCAGACCTACGGCCAGCTGCGCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCAC
CGCCTGATCAACGTGTTCCCCACCAGCGCCTACCGCAGCGAGGCCCTGCGCATCCTGCAGACCCA
GCAGGAGTTCAACCCCCAGATCACCGACGAGTTCATCAACCGCTACCTGGAGATCCTGACCGGCA
AGCGCAAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCACCGACTACGGCCGCTACCGCACC
AGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTGCACCTTCTACCCCGACGA
GTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAACGACCTGAACAACC
TGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAGATCATCAACTACGTG
AAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATCGCCAAGCTGCTGAGCTGCGA
CGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGCAAGGCCGAGATCCACACCTTCGAGG
CCTACCGCAAGATGAAGACCCTGGAGACCCTGGACATCGAGCAGATGGACCGAGAGACCCTGGAC
AAGCTGGCCTACGTGCTGACCCTGAACACCGAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGA
GTTCGCCGACGGCAGCTTCAGCCAGAAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACA
GCAGCATCTTCGGCAAGGGCTGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAG
CTGTACGAGACCAGCGAGGAGCAGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAG
CAGCAGCAACAAGACCAAGTACATCGACGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGG
TGGCCAAGAGCGTGCGCCAGGCCATCAAGATCGTGAACGCCGCCATCAAGGAGTACGGCGACTTC
GACAACATCGTGATCGAGATGGCCCGCGAGACCAACGAGGACGACGAGAAGAAGGCCATCCAGAA
GATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCTGAAGGCCGCCAACCAGTACAACG
GCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTGGCCACCAAGATCCGCCTG
TGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATCAGCATCCACGACCTGATCAA
CAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTGAGCATCACCTTCGACGACAGCCTGG
CCAACAAGGTGCTGGTGTACGCCACCGCCAACCAGGAGAAGGGCCAGCGCACCCCCTACCAGGCC
CTGGACAGCATGGACGACGCCTGGAGCTTCCGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGAC
CCTGAGCAACAAGAAGAAGGAGTATCTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCA
AGAAGTTCATCGAGCGCAACCTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTG
CAGGAGCACTTCCGCGCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAG
CCAGCTGCGCCGCCACTGGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACG
CCCTGATCATTGCGGCTTCTAGCCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGC

FIG. 27a

TACAGCGAGGACCAGCTGCTGGACATCGAGACCGGCGAGCTGATCAGCGACGACGAGTACAAGGA
GAGCGTGTTCAAGGCCCCCTACCAGCACTTCGTGGACACCCTGAAGAGCAAGGAGTTCGAGGACA
GCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAACCGCAAGATCAGCGACGCCACCATCTAC
GCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGGCCGACGAGACCTACGTGCTGGGCAAGATCAA
GGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAGATCTACAAGAAGGACAAGAGCAAGT
TCCTGATGTACCGCCACGACCCCCAGACCTTCGAGAAGGTGATCGAGCCCATCCTGGAGAACTAC
CCCAACAAGCAGATCAACGAGAAAGGCAAGGAGGTGCCCTGCAACCCCTTCCTGAAGTACAAGGA
GGAGCACGGCTACATCCGCAAGTACAGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGAAGT
ACTACGACAGCAAGCTGGGCAACCACATCGACATCACCCCCAAGGACAGCAACAACAAGGTGGTG
CTGCAGAGCGTGAGCCCCTGGCGCGCCGACGTGTACTTCAACAAGAACACCGGCAAGTACGAGAT
CCTGGGGCTGAAGTACAGCGATATGCAGTTTGAGAAAGGCACAGGCAAGTACAGCATCAGCAAGG
AGCAGTACGAGAACATCAAGGTGCGCGAGGGCGTGGACGAGAACAGCGAGTTCAAGTTCACCCTG
TACAAGAACGACCTTCTGCTGCTGAAGGACAGCGAGAACGGCGAGCAAATCCTGCTGCGCTTCAC
CAGCCGCAACGACACCAGCAAGCACTACGTGGAGCTGAAGCCCTACAACCGCCAGAAGTTCGAGG
GCAGCGAGTACCTGATCAAGAGCCTGGGCACCGTGGTGAAGGGCGGCAGGTGCATCAAGGGCCTG
GGCAAGAGCAACATCAGCATCTACAAGGTGCGCACCGACGTGCTGGGCAACCAGCACATCATCAA
*GAACGAGGGCGACAAGCCCAAGTTGGACTTC*AGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGG
GATCC

FIG. 27b

NLS-St1Cas9 LMD-9/TH1477 Hybrid-NLS AA sequence

SV40 NLS
ST1CAS9 SEQUENCE
LINKERS

MGAPKKKRKVGIHGVPAA*SDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGR*
RLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDD
ASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEA
LRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYP
DEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVAD
IKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQ
KQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLL
TEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAAN
QYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLAN
KVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERN
LVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNL
WKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKIS
DATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENY
PNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVS
PWRADVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKD
SENGEQILLRFTSRNDTSKHYVELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVL
*GNQHIIKNEGDKPKLDF*SRADPKKKRKVGS

FIG. 28

NLS-rAPOBEC1-St1Cas9 LMD-9-NLS-2xUGI-NLS-3xHA ORF DNA sequence
NLS
rAPOBEC1
St1Cas9 ORF
UGI
3xHA

ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTC**TCCTCAGAGACTG**
GGCCTGTCGCCGTCGATCCAACCCTGCGCCGCCGGATTGAACCTCACGAGTTTGAAGTGTTCTTTGACCC
CCGGGAGCTGAGAAAGGAGACATGCCTGCTGTACGAGATCAACTGGGGAGGCAGGCACTCCATCTGGAGG
CACACCTCTCAGAACACAAATAAGCACGTGGAGGTGAACTTCATCGAGAAGTTTACCACAGAGCGGTACT
TCTGCCCCAATACCAGATGTAGCATCACATGGTTTCTGAGCTGGTCCCCTTGCGGAGAGTGTAGCAGGGC
CATCACCGAGTTCCTGTCCAGATATCCACACGTGACACTGTTTATCTACATCGCCAGGCTGTATCACCAC
GCAGACCCAAGGAATAGGCAGGGCCTGCGCGATCTGATCAGCTCCGGCGTGACCATCCAGATCATGACAG
AGCAGGAGTCCGGCTACTGCTGGCGGAACTTCGTGAATTATTCTCCTAGCAACGAGGCCCACTGGCCTAG
GTACCCACACCTGTGGGTGCGCCTGTACGTGCTGGAGCTGTATTGCATCATCCTGGGCCTGCCCCCTTGT
CTGAATATCCTGCGGAGAAAGCAGCCCCAGCTGACCTTCTTTACAATCGCCCTGCAGTCTTGTCACTATC
AGAGGCTGCCACCCCACATCCTGTGGGCCACAGGCCTGAAGTCTGGAGGATCTAGCGGAGGCTCCTCTGG
CAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGCGGCAGC*AGC*
GACCTGGTGCTGGGCCTGGCTATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTGACCGGCG
AGATCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAACCTGGTGCGCCGCACCAA
CCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGCGTGCGCCTGAACCGCCTGTTCGAGGAG
AGCGGCCTGATCACCGACTTCACCAAGATCAGCATCAACCTGAACCCCTACCAGCTGCGCGTGAAGGGCC
TGACCGACGAGCTGAGCAACGAGGAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCGGCATCAG
CTACCTGGACGACGCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAAC
AGCAAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGCCAGCTGC
GCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTGTTCCCCACCAGCGCCTA
CCGCAGCGAGGCCCTGCGCATCCTGCAGACCCAGCAGGAGTTCAACCCCCAGATCACCGACGAGTTCATC
AACCGCTACCTGGAGATCCTGACCGGCAAGCGCAAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCA
CCGACTACGGCCGCTACCGCACCAGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTG
CACCTTCTACCCCGACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAAC
GACCTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAGATCATCA
ACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATCGCCAAGCTGCTGAGCTG
CGACGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGCAAGGCCGAGATCCACACCTTCGAGGCC
TACCGCAAGATGAAGACCCTGGAGACCCTGGACATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGG
CCTACGTGCTGACCCTGAACACCGAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGG
CAGCTTCAGCCAGAAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAG
GGCTGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAGCTGTACGAGACCAGCGAGGAGC
AGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAACAAGACCAAGTACATCGA
CGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCCAAGAGCGTGCGCCAGGCCATCAAGATC
GTGAACGCCGCCATCAAGGAGTACGGCGACTTCGACAACATCGTGATCGAGATGGCCCGCGAGACCAACG
AGGACGACGAGAAGAAGGCCATCCAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCT

FIG. 29a

GAAGGCCGCCAACCAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTG
GCCACCAAGATCCGCCTGTGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATCAGCATCC
ACGACCTGATCAACAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTGAGCATCACCTTCGACGA
CAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAGGAGAAGGGCCAGCGCACCCCCTACCAG
GCCCTGGACAGCATGGACGACGCCTGGAGCTTCCGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGACCC
TGAGCAACAAGAAGAAGGAGTATCTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCAAGAAGTT
CATCGAGCGCAACCTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTC
CGCGCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGCCGCCACT
GGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTGATCATTGCGGCTTCTAG
CCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGCTACAGCGAGGACCAGCTGCTGGACATC
GAGACCGGCGAGCTGATCAGCGACGACGAGTACAAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCG
TGGACACCCTGAAGAGCAAGGAGTTCGAGGACAGCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAA
CCGCAAGATCAGCGACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGGCCGACGAG
ACCTACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAGATCTACA
AGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAGAAGGTGATCGAGCCCAT
CCTGGAGAACTACCCCAACAAGCAGATCAACGAGAAAGGCAAGGAGGTGCCCTGCAACCCCTTCCTGAAG
TACAAGGAGGAGCACGGCTACATCCGCAAGTACAGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGA
AGTACTACGACAGCAAGCTGGGCAACCACATCGACATCACCCCCAAGGACAGCAACAACAAGGTGGTGCT
GCAGAGCGTGAGCCCCTGGCGCGCCGACGTGTACTTCAACAAGACCACCGGCAAGTACGAGATCCTGGGG
CTGAAGTACGCCGATCTGCAGTTTGAGAAAGGCACAGGCACCTACAAGATCAGCCAGGAGAAGTACAACG
ACATCAAGAAGAAGGAGGGCGTGGACAGCGACAGCGAGTTCAAGTTCACCCTGTACAAGAACGACCTTCT
GCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTCCGCTTCCTGAGCCGCACCATGCCCAAGCAG
AAGCACTACGTGGAGCTGAAGCCCTACGACAAGCAGAAGTTCGAGGGCGGCGAGGCCCTGATCAAGGTGC
TGGGCAACGTGGCCAACAGCGGCCAGTGCAAGAAGGGCCTGGGCAAGAGCAACATCAGCATCTACAAGGT
*GCGCACCGACGTGCTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTC*AGC
AGGGCTGACCCCAAGAAGAAGAGGAAGGTGGGATCTAGCGGCGGGAGCGGCGGGAGCGGGGGGAGC***ACTA***
ATCTGAGCGACATCATTGAGAAGGAGACTGGGAAACAGCTGGTCATTCAGGAGTCCATCCTGATGCTGCC
TGAGGAGGTGGAGGAAGTGATCGGCAACAAGCCAGAGTCTGACATCCTGGTGCACACCGCCTACGACGAG
TCCACAGATGAGAATGTGATGCTGCTGACCTCTGACGCCCCCGAGTATAAGCCTTGGGCCCTGGTCATCC
AGGATTCTAACGGCGAGAATAAGATCAAGATGCTGAGCGGAGGATCCGGAGGATCTGGAGGCAGC***ACCAA***
CCTGTCTGACATCATCGAGAAGGAGACAGGCAAGCAGCTGGTCATCCAGGAGAGCATCCTGATGCTGCCC
GAAGAAGTCGAAGAAGTGATCGGAAACAAGCCTGAGAGCGATATCCTGGTCCATACCGCCTACGACGAGA
GTACCGACGAAAATGTGATGCTGCTGACATCCGACGCCCCAGAGTATAAGCCCTGGGCTCTGGTCATCCA
GGATTCCAACGGAGAGAACAAAATCAAAATGCTGTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGCGAA
TTCGAGCCCAAGAAGAAGAGGAAAGTCGGATCC***TACCCATACGATGTTCCAGATTACGCTTATCCCTACG***
ACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC

FIG. 29b

NLS-rAPOBEC1-St1Cas9 LMD-9-NLS-2xUGI-NLS-3xHA AA sequence

NLS
rAPOBEC1
St1Cas9 ORF
UGI
3xHA

MKRTADGSEFESPKKKRKV**SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWR**
HTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC
LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGS*S*
DLVLGLAIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEE
SGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKEN
SKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFI
NRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLN
DLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEA
YRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGK
GWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKI
VNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQL
ATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ
ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHF
RAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDI
ETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADE
TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLK
YKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILG
LKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQ
KHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDFS
RADPKKKRKVGSSGGSGGSGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDE***
STDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGS***TNLSDIIEKETGKQLVIQESILMLP***
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSKRTADGSE
FEPKKKRKVGS***YPYDVPDYAYPYDVPDYAYPYDVPDYA***

FIG. 30

NLS-rAPOBEC1-St1Cas9 LMD-9/LMG18311-NLS-2xUGI-NLS-3xHA ORF DNA sequence
NLS
rAPOBEC1
St1Cas9 ORF
UGI
3xHA

ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTC**TCCTCAGAGACTG**
GGCCTGTCGCCGTCGATCCAACCCTGCGCCGCCGGATTGAACCTCACGAGTTTGAAGTGTTCTTTGACCC
CCGGGAGCTGAGAAAGGAGACATGCCTGCTGTACGAGATCAACTGGGGAGGCAGGCACTCCATCTGGAGG
CACACCTCTCAGAACACAAATAAGCACGTGGAGGTGAACTTCATCGAGAAGTTTACCACAGAGCGGTACT
TCTGCCCCAATACCAGATGTAGCATCACATGGTTTCTGAGCTGGTCCCCTTGCGGAGAGTGTAGCAGGGC
CATCACCGAGTTCCTGTCCAGATATCCACACGTGACACTGTTTATCTACATCGCCAGGCTGTATCACCAC
GCAGACCCAAGGAATAGGCAGGGCCTGCGCGATCTGATCAGCTCCGGCGTGACCATCCAGATCATGACAG
AGCAGGAGTCCGGCTACTGCTGGCGGAACTTCGTGAATTATTCTCCTAGCAACGAGGCCCACTGGCCTAG
GTACCCACACCTGTGGGTGCGCCTGTACGTGCTGGAGCTGTATTGCATCATCCTGGGCCTGCCCCCTTGT
CTGAATATCCTGCGGAGAAAGCAGCCCCAGCTGACCTTCTTTACAATCGCCCTGCAGTCTTGTCACTATC
AGAGGCTGCCACCCCACATCCTGTGGGCCACAGGCCTGAAGTCTGGAGGATCTAGCGGAGGCTCCTCTGG
CAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGCGGCAGC*AGC*
GACCTGGTGCTGGGCCTGGCTATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTGACCGGCG
AGATCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAACCTGGTGCGCCGCACCAA
CCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGCGTGCGCCTGAACCGCCTGTTCGAGGAG
AGCGGCCTGATCACCGACTTCACCAAGATCAGCATCAACCTGAACCCCTACCAGCTGCGCGTGAAGGGCC
TGACCGACGAGCTGAGCAACGAGGAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCGGCATCAG
CTACCTGGACGACGCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAAC
AGCAAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGCCAGCTGC
GCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTGTTCCCCACCAGCGCCTA
CCGCAGCGAGGCCCTGCGCATCCTGCAGACCCAGCAGGAGTTCAACCCCCAGATCACCGACGAGTTCATC
AACCGCTACCTGGAGATCCTGACCGGCAAGCGCAAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCA
CCGACTACGGCCGCTACCGCACCAGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTG
CACCTTCTACCCCGACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAAC
GACCTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAGATCATCA
ACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATCGCCAAGCTGCTGAGCTG
CGACGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGCAAGGCCGAGATCCACACCTTCGAGGCC
TACCGCAAGATGAAGACCCTGGAGACCCTGGACATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGG
CCTACGTGCTGACCCTGAACACCGAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGG
CAGCTTCAGCCAGAAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAG
GGCTGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAGCTGTACGAGACCAGCGAGGAGC
AGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAACAAGACCAAGTACATCGA
CGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCCAAGAGCGTGCGCCAGGCCATCAAGATC
GTGAACGCCGCCATCAAGGAGTACGGCGACTTCGACAACATCGTGATCGAGATGGCCCGCGAGACCAACG
AGGACGACGAGAAGAAGGCCATCCAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCT
GAAGGCCGCCAACCAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTG

FIG. 31a

GCCACCAAGATCCGCCTGTGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATCAGCATCC
ACGACCTGATCAACAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTGAGCATCACCTTCGACGA
CAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAGGAGAAGGGCCAGCGCACCCCCTACCAG
GCCCTGGACAGCATGGACGACGCCTGGAGCTTCCGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGACCC
TGAGCAACAAGAAGAAGGAGTATCTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCAAGAAGTT
CATCGAGCGCAACCTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTC
CGCGCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGCCGCCACT
GGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTGATCATTGCGGCTTCTAG
CCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGCTACAGCGAGGAGCAGCTGCTGGACATC
GAGACCGGCGAGCTGATCAGCGACGACGAGTACAAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCG
TGGACACCCTGAAGAGCAAGGAGTTCGAGGACAGCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAA
CCGCAAGATCAGCGACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGAAGGACGAG
ACCTACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAGATCTACA
AGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAGAAGGTGATCGAGCCCAT
CCTGGAGAACTACCCCAACAAGCAGATGAACGAGAAAGGCAAGGAGGTGCCCTGCAACCCCTTCCTGAAG
TACAAGGAGGAGCACGGCTACATCCGCAAGTACAGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGA
AGTACTACGACAGCAAGCTGCTGGGCAACCCCATCGACATCACCCCCGAAAACAGCAAGAACAAGGTGGT
GCTGCAGAGCCTTAAGCCCTGGCGCACCGACGTGTACTTCAACAAGAACACCGGCAAGTACGAGATCCTG
GGGCTGAAGTACGCCGATCTGCAGTTTGAGAAAAAGACAGGCACCTACAAGATCAGCCAGGAGAAGTACA
ACGGCATCATGAAGGAGGAGGGCGTGGACAGCGACAGCGAGTTCAAGTTCACCCTGTACAAGAACGACCT
TCTGCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTCCGCTTCCTGAGCCGCACCATGCCCAAC
GTGAAGTACTACGTGGAGCTGAAGCCCTACAGCAAGGACAAGTTCGAGAAGAACGAGAGCCTGATCGAGA
TCCTGGGCAGCGCCGACAAGAGCGGCAGGTGCATCAAGGGCCTGGGCAAGAGCAACATCAGCATCTACAA
GGTGCGCACCGACGTGCTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTC
AGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGGATCTAGCGGCGGGAGCGGCGGGAGCGGGGGGAGC***A***
CTAATCTGAGCGACATCATTGAGAAGGAGACTGGGAAACAGCTGGTCATTCAGGAGTCCATCCTGATGCT
GCCTGAGGAGGTGGAGGAAGTGATCGGCAACAAGCCAGAGTCTGACATCCTGGTGCACACCGCCTACGAC
GAGTCCACAGATGAGAATGTGATGCTGCTGACCTCTGACGCCCCCGAGTATAAGCCTTGGGCCCTGGTCA
TCCAGGATTCTAACGGCGAGAATAAGATCAAGATGCTGAGCGGAGGATCCGGAGGATCTGGAGGCAGC***AC***
CAACCTGTCTGACATCATCGAGAAGGAGACAGGCAAGCAGCTGGTCATCCAGGAGAGCATCCTGATGCTG
CCCGAAGAAGTCGAAGAAGTGATCGGAAACAAGCCTGAGAGCGATATCCTGGTCCATACCGCCTACGACG
AGAGTACCGACGAAAATGTGATGCTGCTGACATCCGACGCCCCAGAGTATAAGCCCTGGGCTCTGGTCAT
CCAGGATTCCAACGGAGAGAACAAAATCAAAATGCTGTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGC
GAATTCGAGCCCAAGAAGAAGAGGAAAGTCGGATCC***TACCCATACGATGTTCCAGATTACGCTTATCCCT***
ACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC

FIG. 31b

NLS-rAPOBEC1-St1Cas9 LMD-9/LMG18311-NLS-2xUGI-NLS-3xHA AA sequence
NLS
rAPOBEC1
St1Cas9 ORF
UGI
3xHA

MKRTADGSEFESPKKKRKV**SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWR**
HTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC
LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGS*S*
DLVLGLAIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEE
SGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKEN
SKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFI
NRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLN
DLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEA
YRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGK
GWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKI
VNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQL
ATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ
ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHF
RAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDI
ETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDE
TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLK
YKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKNTGKYEIL
GLKYADLQFEKKTGTYKISQEKYNGIMKEEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPN
VKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF
SRADPKKKRKVGSSGGSGGSGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD***
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGS***TNLSDIIEKETGKQLVIQESILML***
PEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSKRTADGS
EFEPKKKRKVGS***YPYDVPDYAYPYDVPDYAYPYDVPDYA***

FIG. 32

NLS-rAPOBEC1-St1Cas9 LMD-9/CNRZ1066-NLS-2xUGI-NLS-3xHA ORF DNA sequence
NLS
rAPOBEC1
St1Cas9 ORF
UGI
3xHA

ATGAAACGGACAGCCGACGGAAGCGAGTTCGAGTCACCAAAGAAGAAGCGGAAAGTC**TCCTCAGAGACTG**
GGCCTGTCGCCGTCGATCCAACCCTGCGCCGCCGGATTGAACCTCACGAGTTTGAAGTGTTCTTTGACCC
CCGGGAGCTGAGAAAGGAGACATGCCTGCTGTACGAGATCAACTGGGGAGGCAGGCACTCCATCTGGAGG
CACACCTCTCAGAACACAAATAAGCACGTGGAGGTGAACTTCATCGAGAAGTTTACCACAGAGCGGTACT
TCTGCCCCAATACCAGATGTAGCATCACATGGTTTCTGAGCTGGTCCCCTTGCGGAGAGTGTAGCAGGGC
CATCACCGAGTTCCTGTCCAGATATCCACACGTGACACTGTTTATCTACATCGCCAGGCTGTATCACCAC
GCAGACCCAAGGAATAGGCAGGGCCTGCGCGATCTGATCAGCTCCGGCGTGACCATCCAGATCATGACAG
AGCAGGAGTCCGGCTACTGCTGGCGGAACTTCGTGAATTATTCTCCTAGCAACGAGGCCCACTGGCCTAG
GTACCCACACCTGTGGGTGCGCCTGTACGTGCTGGAGCTGTATTGCATCATCCTGGGCCTGCCCCCTTGT
CTGAATATCCTGCGGAGAAAGCAGCCCCAGCTGACCTTCTTTACAATCGCCCTGCAGTCTTGTCACTATC
AGAGGCTGCCACCCCACATCCTGTGGGCCACAGGCCTGAAGTCTGGAGGATCTAGCGGAGGCTCCTCTGG
CAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGCGGCAGC*AGC*
GACCTGGTGCTGGGCCTGGCTATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTGACCGGCG
AGATCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAACCTGGTGCGCCGCACCAA
CCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGCGTGCGCCTGAACCGCCTGTTCGAGGAG
AGCGGCCTGATCACCGACTTCACCAAGATCAGCATCAACCTGAACCCCTACCAGCTGCGCGTGAAGGGCC
TGACCGACGAGCTGAGCAACGAGGAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCGGCATCAG
CTACCTGGACGACGCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAAC
AGCAAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGCCAGCTGC
GCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTGTTCCCCACCAGCGCCTA
CCGCAGCGAGGCCCTGCGCATCCTGCAGACCCAGCAGGAGTTCAACCCCCAGATCACCGACGAGTTCATC
AACCGCTACCTGGAGATCCTGACCGGCAAGCGCAAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCA
CCGACTACGGCCGCTACCGCACCAGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTG
CACCTTCTACCCCGACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAAC
GACCTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAGATCATCA
ACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATCGCCAAGCTGCTGAGCTG
CGACGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGCAAGGCCGAGATCCACACCTTCGAGGCC
TACCGCAAGATGAAGACCCTGGAGACCCTGGACATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGG
CCTACGTGCTGACCCTGAACACCGAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGG
CAGCTTCAGCCAGAAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAG
GGCTGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAGCTGTACGAGACCAGCGAGGAGC
AGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAACAAGACCAAGTACATCGA
CGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCCAAGAGCGTGCGCCAGGCCATCAAGATC
GTGAACGCCGCCATCAAGGAGTACGGCGACTTCGACAACATCGTGATCGAGATGGCCCGCGAGACCAACG
AGGACGACGAGAAGAAGGCCATCCAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCT
GAAGGCCGCCAACCAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTG
GCCACCAAGATCCGCCTGTGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATCAGCATCC

FIG. 33a

ACGACCTGATCAACAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTGAGCATCACCTTCGACGA
CAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAGGAGAAGGGCCAGCGCACCCCCTACCAG
GCCCTGGACAGCATGGACGACGCCTGGAGCTTCCGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGACCC
TGAGCAACAAGAAGAAGGAGTATCTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCAAGAAGTT
CATCGAGCGCAACCTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTC
CGCGCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGCCGCCACT
GGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTGATCATTGCGGCTTCTAG
CCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGCTACAGCGAGGAgCAGCTGCTGGACATC
GAGACCGGCGAGCTGATCAGCGACGACGAGTACAAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCG
TGGACACCCTGAAGAGCAAGGAGTTCGAGGACAGCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAA
CCGCAAGATCAGCGACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGaagGACGAG
ACCTACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAGATCTACA
AGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAGAAGGTGATCGAGCCCAT
CCTGGAGAACTACCCCAACAAGCAGATgAACGAGAAAGGCAAGGAGGTGCCCTGCAACCCCTTCCTGAAG
TACAAGGAGGAGCACGGCTACATCCGCAAGTACAGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGA
AGTACTACGACAGCAAGCTGctgGGCAACCcCATCGACATCACCCCCgAaaACAGCAAgAACAAGGTGGT
GCTGCAGAGCcTtAagCCCTGGCGCaCCGACGTGTACTTCAACAAGgCCACCGGCAAGTACGAGATCCTG
GGGCTGAAGTACGCCGATCTGCAGTTTGAGAAAGGCACAGGCACCTACAAGATCAGCCAGGAGAAGTACA
ACGACATCAAGAAGAAGGAGGGCGTGGACAGCGACAGCGAGTTCAAGTTCACCCTGTACAAGAACGACCT
TCTGCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTCCGCTTCCTGAGCCGCACCcTGCCCAAG
CAGAAGCACTACGTGGAGCTGAAGCCCTACGACAAGCAGAAGTTCGAGGGCGGCGAGGCCCTGATCAAGG
TGCTGGGCAACGTGGCCAACgGCGGCCAGTGCAtcAAGGGCCTGGcCAAGAGCAACATCAGCATCTACAA
GGTGCGCACCGACGTGCTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTC
AGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGGATCTAGCGGCGGGAGCGGCGGGAGCGGGGGGAGC***A***
CTAATCTGAGCGACATCATTGAGAAGGAGACTGGGAAACAGCTGGTCATTCAGGAGTCCATCCTGATGCT
GCCTGAGGAGGTGGAGGAAGTGATCGGCAACAAGCCAGAGTCTGACATCCTGGTGCACACCGCCTACGAC
GAGTCCACAGATGAGAATGTGATGCTGCTGACCTCTGACGCCCCCGAGTATAAGCCTTGGGCCCTGGTCA
TCCAGGATTCTAACGGCGAGAATAAGATCAAGATGCTGAGCGGAGGATCCGGAGGATCTGGAGGCAGC***AC***
CAACCTGTCTGACATCATCGAGAAGGAGACAGGCAAGCAGCTGGTCATCCAGGAGAGCATCCTGATGCTG
CCCGAAGAAGTCGAAGAAGTGATCGGAAACAAGCCTGAGAGCGATATCCTGGTCCATACCGCCTACGACG
AGAGTACCGACGAAAATGTGATGCTGCTGACATCCGACGCCCCAGAGTATAAGCCCTGGGCTCTGGTCAT
CCAGGATTCCAACGGAGAGAACAAAATCAAAATGCTGTCTGGCGGCTCAAAAAGAACCGCCGACGGCAGC
GAATTCGAGCCCAAGAAGAAGAGGAAAGTCGGATCC***TACCCATACGATGTTCCAGATTACGCTTATCCCT***
ACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCC

FIG. 33b

NLS-rAPOBEC1-St1Cas9 LMD-9/CNRZ1066-NLS-2xUGI-NLS-3xHA AA sequence
NLS
rAPOBEC1
St1Cas9 ORF
UGI
3xHA

MKRTADGSEFESPKKKRKV**SSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWR**
HTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPC
LNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGS*S*
DLVLGLAIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEE
SGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKEN
SKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFI
NRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLN
DLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEA
YRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGK
GWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKI
VNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQL
ATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ
ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHF
RAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDI
ETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDE
TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLK
YKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKATGKYEIL
GLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPK
QKHYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF
SRADPKKKRKVGSSGGSGGSGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD***
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGS***TNLSDIIEKETGKQLVIQESILML***
PEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSKRTADGS
EFEPKKKRKVGS***YPYDVPDYAYPYDVPDYAYPYDVPDYA***

FIG. 34

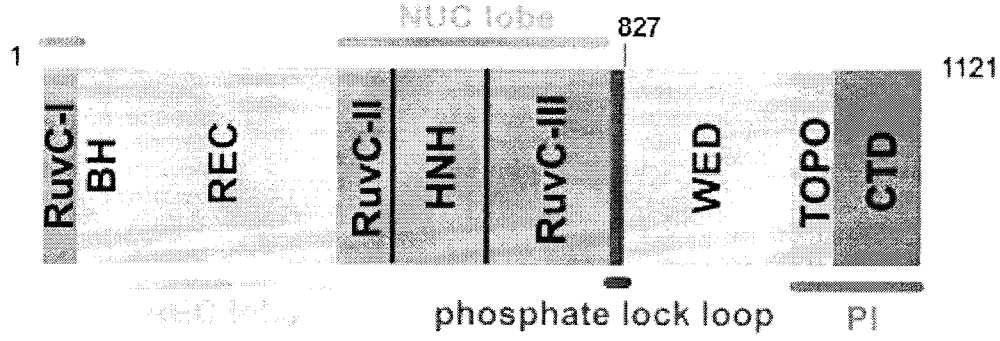


FIG. 35a

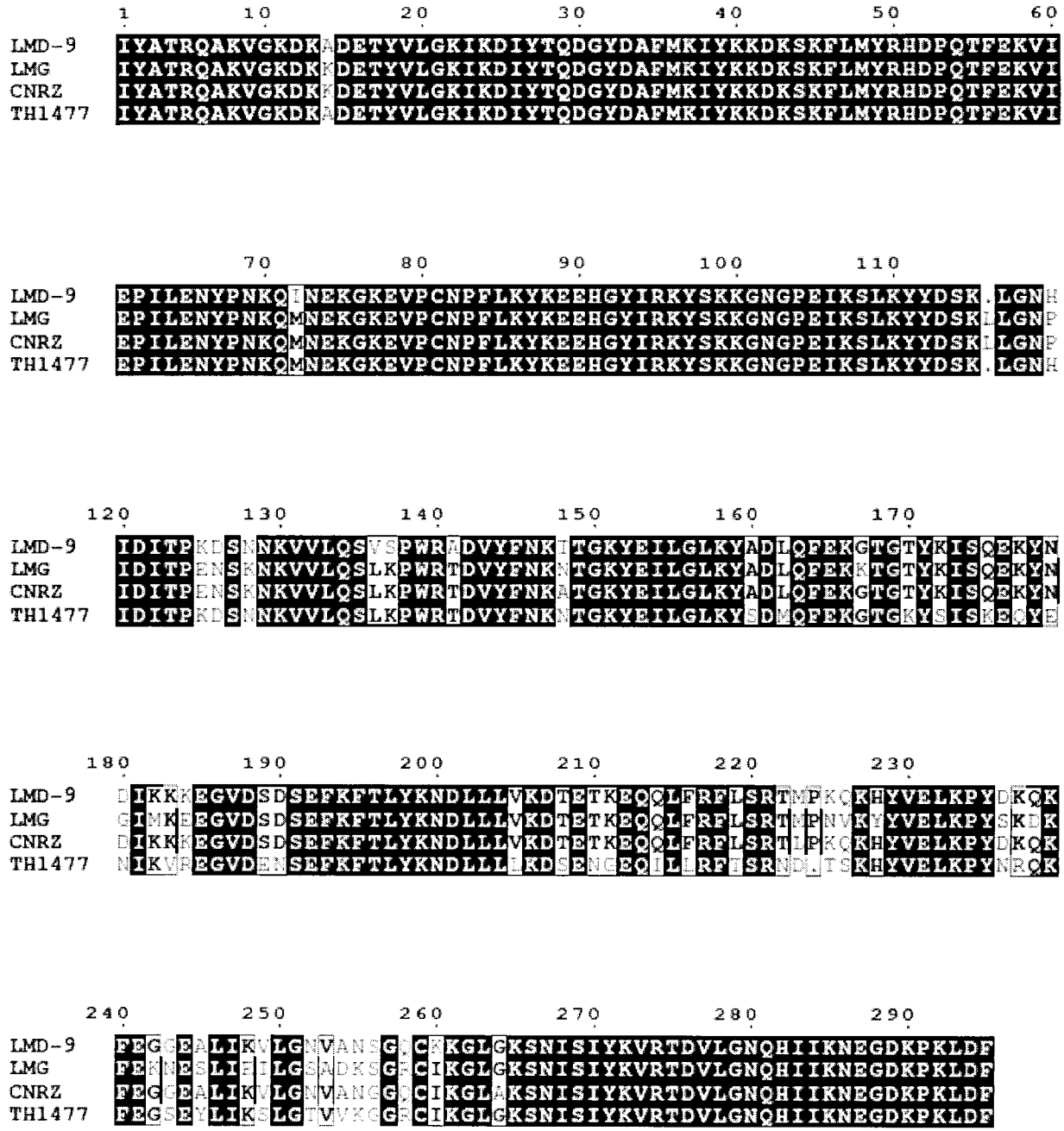


FIG. 35b

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2019/050629

A. CLASSIFICATION OF SUBJECT MATTER
IPC: ***C12N 15/11*** (2006.01), ***A61K 38/47*** (2006.01), ***A61K 48/00*** (2006.01), ***A61P 1/16*** (2006.01), ***A61P 3/00*** (2006.01), ***C12N 15/09*** (2006.01) (more IPCs on the last page)

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)
Keyword search across whole of IPC

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

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

Databases : Google, Google Patent, Google Scholar, NCBI, Questel Orbit, Canadian Patent Database
Keywords : sgRNA, hairpin, RNA, polymerase, termination, stem, loop, modified

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2016/089433 A1 (RYAN, D.E. et al.) 9 June 2016 (09-06-2016) (please see whole document, particularly figure 5B and page 57)	1-6, 8-13 and 32-73
Y		14-29
X	US 2016/0355797 A1 (KONERMANN, S. et al.) 8 December 2016 (08-12-2016) (please see whole document, particularly figure 16A and page 35)	1-13 and 30-73

☐ 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
"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 08 August 2019 (08-08-2019)	Date of mailing of the international search report 09 August 2019 (09-08-2019)
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 819-953-2476	Authorized officer Brad Temple (819) 576-2089

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2019/050629

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of the 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. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim 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. ☐ Claim Nos.:
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:
Please see extra sheet

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 claim 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 claim Nos.:

1-73

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.

C12N 15/55 (2006.01) , ***C12N 15/63*** (2006.01) , ***C12N 15/864*** (2006.01) , ***C12N 15/90*** (2006.01) , ***C12N 9/22*** (2006.01) , *C07H 21/02* (2006.01)

The claims are directed to a plurality of inventive concepts as follows:

Group A - Claims 1-73 are directed to a sgRNA for modification of a target polynucleotide in a cell comprising, (a) a guide segment comprising a guide sequence corresponding to a region of the target polynucleotide, (b) a first hairpin-forming segment located 3' to the guide sequence, the first hairpin-forming segment being capable of forming a hairpin comprising a stem portion and a loop portion, wherein the stem portion does not comprise a sequence corresponding to an RNA polymerase Ill termination signal, and further to related nucleic acids, vectors, host cells, compositions, systems, methods and uses; and

Group B - Claims 74-136 are directed to an isolated CRISPR nuclease polypeptide comprising a first domain and a second domain C-terminal to the first domain, wherein the first domain comprises a guide RNA-binding domain and a nuclease domain, and the second domain comprises a WED domain and a PAM-interacting domain, wherein the first and second domains are derived from different bacterial strains, and further to related nucleic acids, vectors, compositions, systems, methods, and uses.

The claims must be limited to one inventive concept as set out in PCT Rule 13.

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
WO2016089433A1	09 June 2016 (09-06-2016)	WO2016089433A1	09 June 2016 (09-06-2016)
		AU2015355546A1	13 July 2017 (13-07-2017)
		CA2969619A1	09 June 2016 (09-06-2016)
		CN107250148A	13 October 2017 (13-10-2017)
		EP3227447A1	11 October 2017 (11-10-2017)
		EP3227447A4	11 July 2018 (11-07-2018)
		JP2017537626A	21 December 2017 (21-12-2017)
		KR20170084334A	19 July 2017 (19-07-2017)
		US2018051281A1	22 February 2018 (22-02-2018)
		US10337001B2	02 July 2019 (02-07-2019)
		US2016289675A1	06 October 2016 (06-10-2016)
US2016355797A1	08 December 2016 (08-12-2016)	US2016355797A1	08 December 2016 (08-12-2016)
		AU2014361834A1	16 June 2016 (16-06-2016)
		AU2014361834A8	07 July 2016 (07-07-2016)
		CA2932439A1	18 June 2015 (18-06-2015)
		CN105899657A	24 August 2016 (24-08-2016)
		EP3080259A1	19 October 2016 (19-10-2016)
		EP3080260A1	19 October 2016 (19-10-2016)
		EP3080260B1	06 March 2019 (06-03-2019)
		EP3080271A2	19 October 2016 (19-10-2016)
		JP2017503485A	02 February 2017 (02-02-2017)
		JP2017532001A	02 November 2017 (02-11-2017)
		KR20160097327A	17 August 2016 (17-08-2016)
		US2016340660A1	24 November 2016 (24-11-2016)
		US2016355795A1	08 December 2016 (08-12-2016)
		US2018066242A1	08 March 2018 (08-03-2018)
		WO2015089364A1	18 June 2015 (18-06-2015)
		WO2015089364A9	06 August 2015 (06-08-2015)
		WO2015089427A1	18 June 2015 (18-06-2015)
		WO2015089473A1	18 June 2015 (18-06-2015)
		WO2015089473A9	13 August 2015 (13-08-2015)
		WO2015089486A2	18 June 2015 (18-06-2015)
		WO2015089486A3	20 August 2015 (20-08-2015)