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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2019 (28.02.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/040650 A1

(51) International Patent Classification:
A61K 38/00 (2006.01) *C07K 19/00* (2006.01)
C07K 14/195 (2006.01) *C12N 5/10* (2006.01)
C07K 14/315 (2006.01) *C12N 9/14* (2006.01)

(21) International Application Number: PCT/US2018/047577

(22) International Filing Date: 22 August 2018 (22.08.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/549,303 23 August 2017 (23.08.2017) US
62/641,687 12 March 2018 (12.03.2018) US

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, Massachusetts 02114 (US).

(72) Inventors: JOUNG, Keith, J.; 1 Magnolia Way, Winchester, Massachusetts 01890 (US). **KLEINSTIVER, Benjamin**; 93 Fourth Street, Medford, Massachusetts 02155 (US).

(74) Agent: DEYOUNG, Janice Kugler et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
- — *with international search report (Art. 21(3))*
- — *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: ENGINEERED CRISPR-CAS9 NUCLEASES WITH ALTERED PAM SPECIFICITY

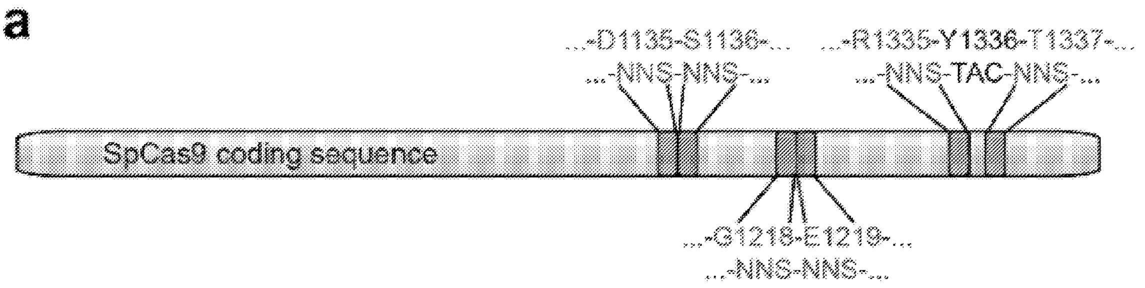


FIG. 1A

(57) Abstract: Engineered CRISPR-Cas9 nucleases with altered and improved PAM specificities and their use in genomic engineering, epigenomic engineering, and genome targeting.

WO 2019/040650 A1

Engineered CRISPR-Cas9 Nucleases with Altered PAM Specificity

CLAIM OF PRIORITY

This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/549,303, filed on August 23, 2017; and 62/641,687, filed on March 12, 2018. The entire contents of the foregoing are hereby incorporated by reference.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with Government support under Grant Nos. GM105378, GM107427, GM118158, and GM088040 awarded by the National Institutes of Health. The Government has certain rights in the invention.

TECHNICAL FIELD

The invention relates, at least in part, to engineered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)/CRISPR-associated protein 9 (Cas9) nucleases with altered and improved Protospacer Adjacent Motif (PAM) specificities and their use in genomic engineering, epigenomic engineering, and genome targeting.

BACKGROUND

CRISPR-Cas9 nucleases enable efficient, customizable genome editing in a wide variety of organisms and cell types (Sander & Joung, Nat Biotechnol 32, 347-355 (2014); Hsuet al., Cell 157, 1262-1278 (2014); Doudna & Charpentier, Science 346, 1258096 (2014); Barrangou & May, Expert Opin Biol Ther 15, 311-314 (2015)). Target site recognition by Cas9 is directed by two short RNAs known as the crRNA and tracrRNA (Deltcheva et al., Nature 471, 602-607 (2011); Jinek et al., Science 337, 816-821 (2012)), which can be fused into a chimeric single guide RNA (**sgRNA**) (Jinek et al., Science 337, 816-821 (2012); Jinek et al., Elife 2, e00471 (2013); Mali et al., Science 339, 823-826 (2013); Cong et al., Science 339, 819-823 (2013)). The 5' end of the sgRNA (derived from the crRNA) can base pair with the target DNA site, thereby permitting straightforward re-programming of site-specific cleavage by the Cas9/sgRNA complex (Jinek et al., Science 337, 816-821 (2012)). However, Cas9 must also recognize a specific protospacer adjacent motif (PAM) that lies proximal to the DNA that base pairs with the sgRNA (Mojica et al., Microbiology 155, 733-740 (2009); Shah et al., RNA Biol 10, 891-899 (2013); Jinek et al., Science 337, 816-821 (2012); Sapranauskas et al, Nucleic Acids Res 39, 9275-9282 (2011); Horvath et al., J Bacteriol 190, 1401-1412 (2008)), a requirement that is needed to initiate sequence-specific recognition (Sternberg et al., Nature 507, 62-67 (2014)) but that can also constrain the targeting range of these nucleases for genome editing. The broadly used *Streptococcus pyogenes* Cas9 (**SpCas9**) recognizes a short NGG PAM (Jinek et al., Science 337, 816-821 (2012); Jiang et al., Nat Biotechnol 31, 233-239 (2013)), which occurs once in every 8 bps of random DNA sequence. By contrast, other Cas9 orthologues characterized to date can recognize longer PAMs (Horvath et al., J Bacteriol 190, 1401-1412 (2008); Fonfara et al., Nucleic Acids Res 42, 2577-2590 (2014); Esvelt et al., Nat Methods 10, 1116-1121 (2013); Ran et al., Nature 520, 186-191 (2015); Zhang et al., Mol Cell 50, 488-503 (2013)). For example, *Staphylococcus aureus* Cas9 (**SaCas9**), one of several smaller Cas9 orthologues that are better suited for viral delivery (Horvath et al., J Bacteriol 190, 1401-1412 (2008); Ran et al., Nature 520, 186-191 (2015); Zhang et al., Mol Cell 50, 488-503 (2013)), recognizes a longer NNGRRT PAM that is expected to occur once in every 32 bps of random DNA. Broadening the targeting range of Cas9 orthologues is important for various applications including the modification of small genetic elements (e.g., transcription factor binding sites (Canver et al. Nature. 527(7577):192-7 (2015); Vierstra et al., Nat Methods. 12(10):927-30 (2015)) or performing allele-specific alterations by positioning sequence differences within the PAM (Courtney, D.G. et al. Gene Ther. 23(1):108-12 (2015). We previously engineered variants of SpCas9 that could recognize sites with NGA and NGCG PAM sequences (Kleinstiver et al, Nature 2015; WO 2016141224), yet many alternative PAM sequences remain untargetable.

SUMMARY

As described herein, the commonly used *Streptococcus pyogenes* Cas9 (**SpCas9**) was engineered to recognize novel PAM sequences using structural information, bacterial selection-based directed evolution, and combinatorial design. These altered PAM specificity variants enable robust editing of reporter sites and endogenous gene sites in human cells that cannot be efficiently targeted by wild-type SpCas9. The present findings provide broadly useful SpCas9 variants, referred to collectively herein as “variants” or “the variants”.

In a first aspect, the invention provides isolated *Streptococcus pyogenes* Cas9 (**SpCas9**) proteins with mutations at three, four, five, or all six of the following positions: D1135, S1136, G1218, E1219, R1335, and/or T1337, e.g., at two, three, four, five, or all of D1135, S1136, G1218, E1219, R1335, and/or T1337 (D1135X/S1136X/G1218X/E1219X/R1335X/T1337X, where X is any amino acid), e.g., comprising a sequence that is at least 80%, 90%, or 95% identical to the amino acid sequence of SEQ ID NO:1 with the indicated mutations. In some embodiments, the variant SpCas9 proteins comprise a set of mutations shown in Tables A, 1, 2, or 3, e.g., one of the following sets of mutations: LRSVQL, LRKIQK, LRSVQK, LWKIQK, VRKIQK, LWKIQK, IRAVQL, VRKLRS, GRKIQK, SWRVVV, SWKVLK, TAHFKV, MSGVKC, LRSVRS, SKTLRP, MWVHLN, TWSMRG, KRRCKV, VRAVQL, VSSVRS, VRSVRS, SRMHCK, GWKLLR, GWKOQK, VAKLLR, VAKIQK, VAKILR, GRKILR, VRKLLR, IRAVQL, VRKIQK, or VRMHCK variant (e.g., for NGTN PAMs); MQKSER, VRKSER, ICKSER, LRSVER, LWLETR, LSRWER, MQSVQL, VRREER, ICCCER, LSRWQR, LWRVVA, WMQAYG, LWRSEY, SQSWRS, LKAWRS, LWGWQH, MCSFER, LWMREQ, LWRVVA, HSSWVR, MWSEPT, GSNYQS, FMQWVN, YCSWVG, MCAWCG, FMQWVR, or SSKWPA variant (e.g., for NGCN PAMs); LRSVRS, SRQMRG, MRARKE, SRMHCK, VRREQR, VRGEQR, LRLSAR, AWTEVTR, KWMMCG, VRGAKE, AWNFQV, LWTTLN, SRMHCK, CWCQCV, AEEQQR, GWEKVR, NRAVNG, LRSYLH, VRGNNR, VQDAQR, GWRQSK, AWLCLS, KWARVV, MWAARP, SRMHCK, VKMAKG, QRKTRE, LCRQQR, CWSHQR, SRTHTQ, LWEVIR, VSSVRS, VRSVRS, IRAVRS, SRSVRS, LWKIQK, or VRMHCK variant (e.g., for NGAN PAMs). In some embodiments, the mutations are not VSREER (also known as VRER) or VSREQR (also known as VRQR).

In some embodiments, the variant SpCas9 proteins comprise one or more mutations that decrease nuclease activity selected from the group consisting of mutations at D10, E762, D839, H983, or D986; and at H840 or N863. In some embodiments, the mutations are: (i) D10A or D10N, and (ii) H840A, H840N, or H840Y.

In some embodiments, the variant SpCas9 proteins comprise one or more mutations that increase specificity selected from the group consisting of mutations at N497, K526, R661, R691, N692, M694, Q695, H698, K810, K848, Q926, K1003, and/or R0160. In some embodiments, the mutations are: N692A, Q695A, Q926A, H698A, N497A, K526A, R661A, R691A, M694A, K810A, K848A, K1003A, R0160A, Y450A/Q695A, L169A/Q695A, Q695A/Q926A, Q695A/D1135E, Q926A/D1135E, Y450A/D1135E, L169A/Y450A/Q695A, L169A/Q695A/Q926A, Y450A/Q695A/Q926A, R661A/Q695A/Q926A, N497A/Q695A/Q926A, Y450A/Q695A/D1135E, Y450A/Q926A/D1135E, Q695A/Q926A/D1135E, L169A/Y450A/Q695A/Q926A, L169A/R661A/Q695A/Q926A,Y450A/R661A/Q695A/Q926A, N497A/Q695A/Q926A/D1135E, R661A/Q695A/Q926A/D1135E, and Y450A/Q695A/Q926A/D1135E; N692A/M694A/Q695A/H698A, N692A/M694A/Q695A/H698A/Q926A; N692A/M694A/Q695A/Q926A; N692A/M694A/H698A/Q926A; N692A/Q695A/H698A/Q926A; M694A/Q695A/H698A/Q926A; N692A/Q695A/H698A; N692A/M694A/Q695A; N692A/H698A/Q926A; N692A/M694A/Q926A; N692A/M694A/H698A; M694A/Q695A/H698A; M694A/Q695A/Q926A; Q695A/H698A/Q926A; G582A/V583A/E584A/D585A/N588A/Q926A; G582A/V583A/E584A/D585A/N588A; T657A/G658A/W659A/R661A/Q926A; T657A/G658A/W659A/R661A; F491A/M495A/T496A/N497A/Q926A; F491A/M495A/T496A/N497A; K918A/V922A/R925A/Q926A; or 918A/V922A/R925A; K855A; K810A/K1003A/R1060A; or K848A/K1003A/R1060A. In some embodiments, the proteins do not include mutations at K526 or R691.

Also provided herein are fusion proteins comprising the isolated variant SpCas9 proteins described herein fused to a heterologous functional domain, with an optional intervening linker, wherein the linker does not interfere with activity of the fusion protein. In some embodiments, the heterologous functional domain is a transcriptional activation domain. In some embodiments, the transcriptional activation domain is from VP64 or NF-κB p65. In some embodiments, the heterologous functional domain is a transcriptional silencer or transcriptional repression domain. In some embodiments, the transcriptional repression domain is a Krueppel-associated box (KRAB) domain, ERF repressor domain (ERD), or mSin3A interaction domain (SID). In some embodiments, the transcriptional silencer is Heterochromatin Protein 1 (HP1), e.g., HP1α or HP1β. In some embodiments, the heterologous functional domain is an enzyme that modifies the methylation state of DNA. In some embodiments, the enzyme that modifies the methylation state of DNA is a DNA methyltransferase (DNMT) or a TET protein. In some embodiments, the TET protein is TET1. In some embodiments, the heterologous functional domain is an enzyme that modifies a histone subunit. In some embodiments, the enzyme that modifies a histone subunit is a histone acetyltransferase (HAT), histone deacetylase (HDAC), histone methyltransferase (HMT), or histone demethylase.

In some embodiments, the heterologous functional domain is a base editor, e.g., a cytidine deaminase domain, e.g., from the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like (APOBEC) family of deaminases, including APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D/E, APOBEC3F, APOBEC3G, APOBEC3H, or APOBEC4; activation-induced cytidine deaminase (AID), e.g., activation induced cytidine deaminase (AICDA); cytosine deaminase 1 (CDA1) or CDA2; or cytosine deaminase acting on tRNA (CDAT). In some embodiments, the heterologous functional domain is a deaminase that modifies adenosine DNA bases, e.g., the deaminase is an adenosine deaminase 1 (ADA1), ADA2; adenosine deaminase acting on RNA 1 (ADAR1), ADAR2, ADAR3; adenosine deaminase acting on tRNA 1 (ADAT1), ADAT2, ADAT3; and naturally occurring or engineered tRNA-specific adenosine deaminase (TadA). In some embodiments, the heterologous functional domain is a biological tether. In some embodiments, the biological tether is MS2, Csy4 or lambda N protein. In some embodiments, the heterologous functional domain is FokI.

In some embodiments, the heterologous functional domain is an enzyme, domain, or peptide that inhibits or enhances endogenous DNA repair or base excision repair (BER) pathways, e.g., uracil DNA glycosylase inhibitor (UGI) that inhibits uracil DNA glycosylase (UDG, also known as uracil N-glycosylase, or UNG) mediated excision of uracil to initiate BER; or DNA end-binding proteins such as Gam from the bacteriophage Mu.

Also provided herein are isolated nucleic acids encoding the variant SpCas9 proteins described herein, as well as vectors comprising the isolated nucleic acids, optionally operably linked to one or more regulatory domains for expressing the variant SpCas9 proteins described herein. Also provided herein are host cells, e.g., mammalian host cells, comprising the nucleic acids described herein, and optionally expressing the variant SpCas9 proteins described herein. Also provided herein are ribonucleoprotein (RNP) complexes that include a variant spCas9 protein as described herein and a guide RNA that targets a sequence having a PAM sequence targeted by the variant protein.

Also provided herein are methods of altering the genome of a cell, by expressing in the cell an isolated variant SpCas9 protein described herein, and a guide RNA having a region complementary to a selected portion of the genome of the cell.

Also provided herein are methods for altering, e.g., selectively altering, the genome of a cell by expressing in the cell the variant proteins, and a guide RNA having a region complementary to a selected portion of the genome of the cell.

Also provided are methods for altering, e.g., selectively altering, the genome of a cell by contacting the cell with a protein variant described herein, and a guide RNA having a region complementary to a selected portion of the genome of the cell.

In some embodiments, the isolated protein or fusion protein comprises one or more of a nuclear localization sequence, cell penetrating peptide sequence, and/or affinity tag.

In some embodiments of the methods described herein, the cell is a stem cell, e.g., an embryonic stem cell, mesenchymal stem cell, or induced pluripotent stem cell; is in a living animal; or is in an embryo, e.g., a mammalian, insect, or fish (e.g., zebrafish) embryo or embryonic cell.

Further, provided herein are methods, e.g., *in vitro* methods, for altering a double stranded DNA (dsDNA) molecule. The methods include contacting the dsDNA molecule with one or more of the variant proteins described herein, and a guide RNA having a region complementary to a selected portion of the dsDNA molecule.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, sequences, database entries, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

Other features and advantages of the invention will be apparent from the following detailed description and FIGs, and from the claims.

DESCRIPTION OF DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

Figures 1A-B. Generation and screening of SpCas9 PAM variant libraries. (a) To generate a library of SpCas9 variants with diverse amino acid combinations at six positions important for PAM recognition, degenerate oligo libraries were cloned into three regions of the SpCas9 coding sequence to randomize the six codons highlighted in red. (b) Schematic of the bacterial selection assay, where SpCas9 variants and an sgRNA are expressed from one plasmid, and a second plasmid encodes an inducible toxic gene and a customizable target site. Both plasmids are co-transformed into bacteria, and only cells that harbor an SpCas9 PAM variant that can recognize the PAM encoded in the target site of the toxic plasmid can survive on media that induces expression of the toxic gene (selective media contains 10 mM arabinose).

Figures 2A-F. SpCas9 PAM variants obtained from bacterial selection that recognize sites with NGTN PAMs. (A-F) Plasmids encoding SpCas9 variants from initial screens against positive selection plasmids carrying NGTG, NGTT, NGTC, or NGTA PAM sequences were re-screened in the bacterial screen assay to assess activity on target sites with NGTN PAM sequences. For panels C, E, and F, variants were also screened against target sites with NGAG, NGAT, NGAC, and NGAA PAMs. Approximate activities of the variants on sites with the indicated PAMs were calculated by comparing the number of colonies on non-selective media (chloramphenicol only) to those on selective media (chloramphenicol + 10 mM arabinose), and calculating the % survival (Table 1). Variants whose numbers have an asterisk indicate plasmids that were tested in prior assays (and shown in above panels of this figure), and the amino acid sequences of these variants can be found in Table 1.

Figures 3A-C. Activities of SpCas9 NGTN PAM variants in human cells. (A,B) Human cell EGFP-disruption assays to assess the activities of various SpCas9 variants (that were obtained through selections or rationally designed) against sites in

EGFP that harbor NGTA, NGTC, NGTT, and NGTG PAMs. Error bars represent s.e.m. for n = 2-7. (C) Endogenous gene disruption activity of one of the top performing SpCas9-NGTN PAM variants, the SpCas9-LRSVQL variant, against sites harboring NGTG, NGTA, NGTC, or NGTT PAMs. Gene disruption quantified by T7E1 assay.

Figures 4A-B. SpCas9 PAM variants obtained from bacterial selection that recognize sites with NGCN PAMs. (A-B) Plasmids encoding SpCas9 variants from initial screens against positive selection plasmids carrying NGCG, NGCT, NGCC, or NGCA PAM sequences were re-screened in the bacterial screen assay to assess activity on target sites with NGCN PAM sequences. Approximate activities of the variants on sites with the indicated PAMs were calculated by comparing the number of colonies on non-selective media (chloramphenicol only) to those on selective media (chloramphenicol + 10 mM arabinose), and calculating the % survival (Table 2). Variants whose numbers have an asterisk indicate plasmids that were tested in prior assays (and shown in above panels of this figure), and the amino acid sequences of these variants can be found in Table 2.

Figures 5A-E. Activities of SpCas9 NGCN PAM variants in human cells. (A-E) Human cell EGFP-disruption assays to assess the activities of various SpCas9 variants (that were obtained through selections or rationally designed) against sites in EGFP that harbor NGCA, NGCC, NGCT, and NGCG PAMs. Activities on sites bearing (A) NGCA PAM sites, (B) NGCC PAM sites, (C) NGCT PAM sites, (D) NGCC and NGCT PAM sites, and (E) various NGCN PAM sites.

Figures 6A-B. SpCas9 PAM variants obtained from bacterial selection that recognize sites with NGAN PAMs. (A-B) Plasmids encoding SpCas9 variants from initial screens against positive selection plasmids carrying NGAG, NGAT, NGAC, or NGAA PAM sequences were re-screened in the bacterial screen assay to assess activity on target sites with NGAN PAM sequences. Approximate activities of the variants on sites with the indicated PAMs were calculated by comparing the number of colonies on non-selective media (chloramphenicol only) to those on selective media (chloramphenicol + 10 mM arabinose), and calculating the % survival (Table 3). The amino acid sequences of these variants can be found in Table 3.

Figures 7A-B. Activities of SpCas9 NGAN PAM variants in human cells. (A-B) Human cell EGFP-disruption assays to assess the activities of various SpCas9 variants (that were obtained through selections or rationally designed) on sites in EGFP that harbor (A) NGAA, NGAC, NGAT, and NGAG PAMs, or (B) various NGAC PAMs.

Figure 8. Comparison of the nuclease activities of SpCas9 PAM variants to xCas9 in human cells. Endogenous gene disruption activities were assessed by T7E1 assay.

Figure 9. Comparison of the base editing activities of SpCas9 PAM variants to xCas9 on endogenous sites in human cells. (A) BE3 versions (Komor et al., Nature. 2016 May 19;533(7603):420-4) of SpCas9 PAM variants were tested for C-to-T base editing activity, as assessed by targeted deep sequencing and analysis performed with CRISPResso (Pinello et al., Nat Biotechnol. 2016 Jul 12; 34(7): 695–697). Nat Biotechnol. 2016 Jul 12; 34(7): 695–697). Edited cytosine nucleotides labeled on the x-axis are numbered starting at position 1 as the most PAM-distal position in the Cas9 target sequence. (B) ABE7.10 versions (Gaudelli et al., Nature. 2017 Nov 23;551(7681):464-471) of SpCas9 PAM variants were tested for A-to-G base editing activity, as assessed by targeted deep sequencing and analysis performed with CRISPResso. Edited adenosine nucleotides labeled on the x-axis are numbered starting at position 1 as the most PAM-distal position in the Cas9 target sequence.

Table 1: Selection results and activity in bacteria of variants on sites harboring NGTN PAMs

Table 2: Selection results and activity in bacteria of variants on sites harboring NGCN PAMs

Table 3: Selection results and activity in bacteria of variants on sites harboring NGAN PAMs

DETAILED DESCRIPTION

Recognition of a protospacer adjacent motif (PAM) by Streptococcus pyogenes Cas9 (SpCas9) is the critical first step of target DNA recognition, enabling SpCas9 to bind and hydrolyze DNA. Although CRISPR-Cas9 nucleases are widely used for genome editing[1-4], the range of sequences that Cas9 can cleave is constrained by the need for a specific PAM in the target site[5, 6]. For example, SpCas9, the most robust and widely used Cas9 to date, primarily recognizes NGG PAMs. As a result, it can often be difficult to target double-stranded breaks (DSBs) with the precision that is necessary for various genome editing applications. In addition, imperfect PAM recognition by Cas9 can lead to the creation of unwanted off-target mutations[7, 8]. Cas9 derivatives with purposefully altered and/or improved PAM specificities would address these limitations.

Crystal structures reveal that wild-type SpCas9 utilizes two arginine amino acid side chains (R1333 and R1335) to make base specific contacts to the guanines of the canonical NGG PAM sequence. However, to alter PAM recognition and improve the targeting range of SpCas9, we and others have shown that simply mutating either one or both of these arginines does not confer a switch in PAM preference (Anders et al, Nature 2014; Kleinstiver et al, Nature 2015; WO 2016141224). We previously undertook a selection approach to evolve variants of SpCas9 that could target NGA and NGCG PAM sequences (Kleinstiver et al, Nature 2015; WO 2016141224); however, many alternative PAM sequences remain untargetable.

To further expand the utility of SpCas9 by enabling targeting of currently inaccessible PAM sequences, we conceived of an alternative strategy to select for SpCas9 variants capable of recognizing novel PAM sequences. Having established previously that certain positions within the SpCas9 coding sequence are important for PAM recognition (Kleinstiver et al., Nature 2015; WO 2016141224), we conducted a focused saturation mutagenesis approach where we randomized six amino acids within three separate regions of the PAM interacting domain to generate a library of SpCas9 variants with diverse codon usage at these positions: D1135/S1136, G1218/E1219, and R1335/T1337. To do so, we sequentially cloned randomized oligonucleotide cassettes encoding NNS nucleotide triplets (where N is any nucleotide and S is G or C) at the codons of SpCas9 that contain encode these six amino acids (Figure 1A).The resulting library of SpCas9 variants was then subjected to selection using our bacterial positive selection assay as previously described (Kleinstiver et al., Nature 2015) to identify variants that can cleave target sites harboring various NGNN PAM sequences (Figure 1B). Briefly, bacteria can only survive selective conditions (plating on 10 mM arabinose, which induces transcription of the ccdB toxic gene) if an expressed SpCas9 variant can recognize the target site (PAM and spacer sequence) encoded in the positive selection plasmid. Strong PAM recognition will lead to hydrolysis of the selection plasmid, preventing induction of *ccdB* expression and thereby allowing bacterial growth. Thus, while screening SpCas9 libraries, colonies that grow on media containing 10 mM arabinose are expected to encode an SpCas9 PAM variant that can target a site bearing an alternate PAM of interest (Figure 1B).

Engineered Cas9 Variants with Altered PAM Specificities

The SpCas9 variants engineered in this study greatly increase the range of target sites accessible by wild-type SpCas9, further enhancing the opportunities to use the CRISPR-Cas9 platform to practice efficient HDR, to target NHEJ-mediated indels to small genetic elements, and to exploit the requirement for a PAM to distinguish between two different alleles in the same cell. The selection and rational design of variants that can now target formerly inaccessible NGTN and NGCH (where H is A, C, or T) PAM containing sites, and variants that can improve activity against NGAC, improve the prospects for accurate and high-resolution genome-editing. The altered PAM specificity SpCas9 variants can efficiently disrupt endogenous gene sites that are not currently targetable by SpCas9 in both bacterial and human cells, suggesting that they will work in a variety of different cell types and organisms.

All of the SpCas9 variants described herein can be rapidly incorporated into existing and widely used vectors, e.g., by simple site-directed mutagenesis, and because they require only a small number of mutations contained within the PAM-interacting domain, the variants should also work with other previously described improvements to the SpCas9 platform (e.g., truncated sgRNAs (Tsai et al., Nat Biotechnol 33, 187-197 (2015); Fu et al., Nat Biotechnol 32, 279-284 (2014)), nickase mutations (Mali et al., Nat Biotechnol 31, 833-838 (2013); Ran et al., Cell 154, 1380-1389 (2013)), dimeric FokI-dCas9 fusions (Guilinger et al., Nat Biotechnol 32, 577-582 (2014); Tsai et al., Nat Biotechnol 32, 569-576 (2014)); and high-fidelity variants (Kleinstiver et al. Nature 2016).

SpCas9 Variants with Altered PAM Specificity

Thus, provided herein are SpCas9 variants. The SpCas9 wild type sequence is as follows:

```
        10         20         30         40         50         60
MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE

        70         80         90        100        110        120
ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG

       130        140        150        160        170        180
NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD
```

```
       190        200        210        220        230        240
VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN

       250        260        270        280        290        300
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI

       310        320        330        340        350        360
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA

       370        380        390        400        410        420
GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH

       430        440        450        460        470        480
AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE

       490        500        510        520        530        540
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL

       550        560        570        580        590        600
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI

       610        620        630        640        650        660
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG

       670        680        690        700        710        720
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL

       730        740        750        760        770        780
HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER

       790        800        810        820        830        840
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDH

       850        860        870        880        890        900
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL

       910        920        930        940        950        960
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS

       970        980        990       1000       1010       1020
KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK

      1030       1040       1050       1060       1070       1080
MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF

      1090       1100       1110       1120       1130       1140
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA

      1150       1160       1170       1180       1190       1200
YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLIIKLPK

      1210       1220       1230       1240       1250       1260
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE

      1270       1280       1290       1300       1310       1320
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA

      1330       1340       1350       1360
PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD  (SEQ ID NO:1)
```

The SpCas9 variants described herein can include mutations at one, two, three, four, five, or all six of the following positions: D1135, S1136, G1218, E1219, R1335, and/or T1337, e.g., D1135X/S1136X/G1218X/E1219X/R1335X/T1337X, where X is any amino acid (or at positions analogous thereto). In some embodiments, the SpCas9 variants are at least 80%, e.g., at least 85%, 90%, or 95% identical to the amino acid sequence of SEQ ID NO:1, e.g., have differences at up to 5%, 10%, 15%, or 20% of the residues of SEQ ID NO:1 replaced, e.g., with conservative mutations. In preferred embodiments, the variant retains desired activity of the parent, e.g., the nuclease activity (except where the parent is a nickase or a dead Cas9), and/or the ability to interact with a guide RNA and target DNA).

To determine the percent identity of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 80% of the length of the reference sequence, and in some embodiments is at least 90% or 100%. The nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein nucleic acid "identity" is equivalent to nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. Percent identity between two polypeptides or nucleic acid sequences is determined in various ways that are within the skill in the art, for instance, using publicly available computer software such as Smith Waterman Alignment (Smith, T. F. and M. S. Waterman (1981) J Mol Biol 147:195-7); "BestFit" (Smith and Waterman, Advances in Applied Mathematics, 482-489 (1981)) as incorporated into GeneMatcher Plus^{TM}, Schwarz and Dayhof (1979) Atlas of Protein Sequence and Structure, Dayhof, M.O., Ed, pp 353-358; BLAST program (Basic Local Alignment Search Tool; (Altschul, S. F., W. Gish, et al. (1990) J Mol Biol 215: 403-10), BLAST-2, BLAST-P, BLAST-N, BLAST-X, WU-BLAST-2, ALIGN, ALIGN-2, CLUSTAL, or Megalign (DNASTAR) software. In addition, those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the length of the sequences being compared. In general, for proteins or nucleic acids, the length of comparison can be any length, up to and including full length (e.g., 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100%). For purposes of the present compositions and methods, at least 80% of the full length of the sequence is aligned using the BLAST algorithm and the default parameters.

For purposes of the present invention, the comparison of sequences and determination of percent identity between two sequences can be accomplished using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

In some embodiments, the SpCas9 variants include a set of mutations shown in Tables 1, 2, or 3, e.g., one of the following sets of mutations at D1135X/S1136X/G1218X/E1219X/R1335X/T1337X: SpCas9-LWKIQK, LWKIQK, IRAVQL, SWRVVV, SWKVLK, TAHFKV, MSGVKC, LRSVRS, SKTLRP, MWVHLN, TWSMRG, KRRCKV, VRAVQL, VSSVRS, VRSVRS, SRMHCK, GRKIQK, GWKLLR, GWKOQK, VAKLLR, VAKIQK, VAKILR, GRKILR, VRKLLR, LRSVQL, IRAVQL, VRKIQK, VRMHCK, LRKIQK, LRSVQK, or VRKIQK variant (e.g., for NGTN PAMs); WMQAYG, MQKSER, LWRSEY, SQSWRS, LKAWRS, LWGWQH, MCSFER, LWMREQ, LWRVVA, HSSWVR, MWSEPT, GSNYQS, FMQWVN, YCSWVG, MCAWCG, LWLETR, FMQWVR, SSKWPA, LSRWQR, ICCCER, VRKSER, or ICKSER (e.g., for NGCN PAMs); or LRLSAR, AWTEVTR, KWMMCG, VRGAKE, MRARKE, AWNFQV, LWTTLN, SRMHCK, CWCQCV, AEEQQR, GWEKVR, NRAVNG, SRQMRG, LRSYLH, VRGNNR, VQDAQR, GWRQSK, AWLCLS, KWARVV, MWAARP, SRMHCK, VKMAKG, QRKTRE, LCRQQR, CWSHQR, SRTHTQ, LWEVIR, VSSVRS, VRSVRS, LRSVRS, IRAVRS, SRSVRS, LWKIQK, VRMHCK, or SRMHCK (e.g., for NGAN PAMs). In some embodiments, the spCas9 variants include D1135L/S1136R/G1218S/E1219V/R1335X/T1337X, e.g., LRSVQL or LRSVRS. In some embodiments, the residue at D1135 is an L, G, I, V, M, or S. In some embodiments, the residue at S1136 is an R, Q, W, S, or C. In some embodiments, the residue at G1218 is an S, K, S, R, L, C, G, A, or Q. In some embodiments, the residue at E1219 is V, I, S, E, W, C, A, or R. In some embodiments, the residue at R1335 is an R, Q, E, V, T, or K. In some embodiments, the residue at T1337 is an S, K, L, R, A, E, T, or G. In some embodiments, the variants include one of the sets of mutations in Table A.

Table A

NGTN PAM	**NGCN PAM**	**NGAN PAM**
LRSVRS	**MQKSER**	VRGAKE
GRKIQK	LWRVVA	MRARKE
LRSVQL	LWLETR	SRQMRG
IRAVQL	LSRWQR	**LRSVRS**
LRKIQK	ICCCER	
LRSVQK	VRKSER	
VRKIQK	ICKSER	

In some embodiments, the SpCas9 variants also include mutations at one of the following amino acid positions, which reduce or destroy the nuclease activity of the Cas9: D10, E762, D839, H983, or D986 and H840 or N863, e.g., D10A/D10N and H840A/H840N/H840Y, to render the nuclease portion of the protein catalytically inactive; substitutions at these positions could be alanine (as they are in Nishimasu al., Cell 156, 935–949 (2014)), or other residues, e.g., glutamine, asparagine, tyrosine, serine, or aspartate, e.g., E762Q, H983N, H983Y, D986N, N863D, N863S, or N863H (see WO 2014/152432). In some embodiments, the variant includes mutations at D10A or H840A (which creates a single-strand nickase), or mutations at D10A and H840A (which abrogates nuclease activity; this mutant is known as dead Cas9 or dCas9).

In some embodiments, the SpCas9 variants also include mutations at one or more amino acid positions that increase the specificity of the protein (i.e., reduce off-target effects). In some embodiments, the SpCas9 variants include one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, or all thirteen mutations at the following residues: N497, K526, R661, R691, N692, M694, Q695, H698, K810, K848, Q926, K1003, and/or R0160. In some embodiments, the mutations are: N692A, Q695A, Q926A, H698A, N497A, K526A, R661A, R691A, M694A, K810A, K848A, K1003A, R0160A, Y450A/Q695A, L169A/Q695A, Q695A/Q926A, Q695A/D1135E, Q926A/D1135E, Y450A/D1135E, L169A/Y450A/Q695A,

L169A/Q695A/Q926A, Y450A/Q695A/Q926A, R661A/Q695A/Q926A, N497A/Q695A/Q926A, Y450A/Q695A/D1135E, Y450A/Q926A/D1135E, Q695A/Q926A/D1135E, L169A/Y450A/Q695A/Q926A, L169A/R661A/Q695A/Q926A,Y450A/R661A/Q695A/Q926A, N497A/Q695A/Q926A/D1135E, R661A/Q695A/Q926A/D1135E, and Y450A/Q695A/Q926A/D1135E; N692A/M694A/Q695A/H698A, N692A/M694A/Q695A/H698A/Q926A; N692A/M694A/Q695A/Q926A; N692A/M694A/H698A/Q926A; N692A/Q695A/H698A/Q926A; M694A/Q695A/H698A/Q926A; N692A/Q695A/H698A; N692A/M694A/Q695A; N692A/H698A/Q926A; N692A/M694A/Q926A; N692A/M694A/H698A; M694A/Q695A/H698A; M694A/Q695A/Q926A; Q695A/H698A/Q926A; G582A/V583A/E584A/D585A/N588A/Q926A; G582A/V583A/E584A/D585A/N588A; T657A/G658A/W659A/R661A/Q926A; T657A/G658A/W659A/R661A; F491A/M495A/T496A/N497A/Q926A; F491A/M495A/T496A/N497A; K918A/V922A/R925A/Q926A; or 918A/V922A/R925A; K855A; K810A/K1003A/R1060A; or K848A/K1003A/R1060A. See, e.g., US9512446B1; Kleinstiver et al., Nature. 2016 Jan 28;529(7587):490-5; Slaymaker et al., Science. 2016 Jan 1;351(6268):84-8; Chen et al., Nature. 2017 Oct 19;550(7676):407-410; Tsai and Joung, Nature Reviews Genetics 17:300–312 (2016); Vakulskas et al., Nature Medicine 24:1216–1224 (2018); Casini et al., Nat Biotechnol. 2018 Mar;36(3):265-271. In some embodiments, the variants do not include mutations at K526 or R691.

In some embodiments, the SpCas9 variants include mutations at one, two, three, four, five, six or all seven of the following positions: L169A, Y450, N497, R661, Q695, Q926, and/or D1135E, e.g., in some embodiments, the variant SpCas9 proteins comprise mutations at one, two, three, or all four of the following: N497, R661, Q695, and Q926, e.g., one, two, three, or all four of the following mutations: N497A, R661A, Q695A, and Q926A. In some embodiments, the variant SpCas9 proteins comprise mutations at Q695 and/or Q926, and optionally one, two, three, four or all five of L169, Y450, N497, R661 and D1135E, e.g., including but not limited to Y450A/Q695A, L169A/Q695A, Q695A/Q926A, Q695A/D1135E, Q926A/D1135E, Y450A/D1135E, L169A/Y450A/Q695A, L169A/Q695A/Q926A, Y450A/Q695A/Q926A, R661A/Q695A/Q926A, N497A/Q695A/Q926A,

Y450A/Q695A/D1135E, Y450A/Q926A/D1135E, Q695A/Q926A/D1135E, L169A/Y450A/Q695A/Q926A, L169A/R661A/Q695A/Q926A,Y450A/R661A/Q695A/Q926A, N497A/Q695A/Q926A/D1135E, R661A/Q695A/Q926A/D1135E, and Y450A/Q695A/Q926A/D1135E. See, e.g., Kleinstiver et al., Nature 529:490–495 (2016); WO 2017/040348; US 9,512,446).

In some embodiments, the SpCas9 variants also include mutations at one, two, three, four, five, six, seven, or more of the following positions: F491, M495, T496, N497, G582, V583, E584, D585, N588, T657, G658, W659, R661, N692, M694, Q695, H698, K918, V922, and/or R925, and optionally at Q926, preferably comprising a sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO:1 with mutations at one, two, three, four, five, six, seven, or more of the following positions: F491, M495, T496, N497, G582, V583, E584, D585, N588, T657, G658, W659, R661, N692, M694, Q695, H698, K918, V922, and/or R925, and optionally at Q926, and optionally one or more of a nuclear localization sequence, cell penetrating peptide sequence, and/or affinity tag.

In some embodiments, the proteins comprise mutations at one, two, three, or all four of the following: N692, M694, Q695, and H698; G582, V583, E584, D585, and N588; T657, G658, W659, and R661; F491, M495, T496, and N497; or K918, V922, R925, and Q926.

In some embodiments, the proteins comprise one, two, three, four, or all of the following mutations: N692A, M694A, Q695A, and H698A; G582A, V583A, E584A, D585A, and N588A; T657A, G658A, W659A, and R661A; F491A, M495A, T496A, and N497A; or K918A, V922A, R925A, and Q926A.

In some embodiments, the proteins comprise mutations: N692A/M694A/Q695A/H698A.

In some embodiments, the proteins comprise mutations: N692A/M694A/Q695A/H698A/Q926A; N692A/M694A/Q695A/Q926A; N692A/M694A/H698A/Q926A; N692A/Q695A/H698A/Q926A; M694A/Q695A/H698A/Q926A; N692A/Q695A/H698A; N692A/M694A/Q695A; N692A/H698A/Q926A; N692A/M694A/Q926A; N692A/M694A/H698A; M694A/Q695A/H698A; M694A/Q695A/Q926A; Q695A/H698A/Q926A; G582A/V583A/E584A/D585A/N588A/Q926A;

G582A/V583A/E584A/D585A/N588A; T657A/G658A/W659A/R661A/Q926A; T657A/G658A/W659A/R661A; F491A/M495A/T496A/N497A/Q926A; F491A/M495A/T496A/N497A; K918A/V922A/R925A/Q926A; or 918A/V922A/R925A. See, e.g., Chen et al., "Enhanced proofreading governs CRISPR-Cas9 targeting accuracy," bioRxiv, doi.org/10.1101/160036 (August 12, 2017).

In some embodiments, the variant proteins include mutations at one or more of R780, K810, R832, K848, K855, K968, R976, H982, K1003, K1014, K1047, and/or R1060, e.g., R780A, K810A, R832A, K848A, K855A, K968A, R976A, H982A, K1003A, K1014A, K1047A, and/or R1060A, e.g., K855A; K810A/K1003A/R1060A; (also referred to as eSpCas9 1.0); or K848A/K1003A/R1060A (also referred to as eSpCas9 1.1) (see Slaymaker et al., Science. 2016 Jan 1;351(6268):84-8).

Also provided herein are isolated nucleic acids encoding the SpCas9 variants, vectors comprising the isolated nucleic acids, optionally operably linked to one or more regulatory domains for expressing the variant proteins, and host cells, e.g., mammalian host cells, comprising the nucleic acids, and optionally expressing the variant proteins.

The variants described herein can be used for altering the genome of a cell; the methods generally include expressing the variant proteins in the cells, along with a guide RNA having a region complementary to a selected portion of the genome of the cell. Methods for selectively altering the genome of a cell are known in the art, see, e.g., US8,697,359; US2010/0076057; US2011/0189776; US2011/0223638; US2013/0130248; WO/2008/108989; WO/2010/054108; WO/2012/164565; WO/2013/098244; WO/2013/176772; US20150050699; US20150045546; US20150031134; US20150024500; US20140377868; US20140357530; US20140349400; US20140335620; US20140335063; US20140315985; US20140310830; US20140310828; US20140309487; US20140304853; US20140298547; US20140295556; US20140294773; US20140287938; US20140273234; US20140273232; US20140273231; US20140273230; US20140271987; US20140256046; US20140248702; US20140242702; US20140242700; US20140242699; US20140242664; US20140234972; US20140227787; US20140212869; US20140201857; US20140199767; US20140189896; US20140186958; US20140186919; US20140186843;

US20140179770; US20140179006; US20140170753; Makarova et al., "Evolution and classification of the CRISPR-Cas systems" 9(6) Nature Reviews Microbiology 467-477 (1-23) (Jun. 2011); Wiedenheft et al., "RNA-guided genetic silencing systems in bacteria and archaea" 482 Nature 331-338 (Feb. 16, 2012); Gasiunas et al., "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria" 109(39) Proceedings of the National Academy of Sciences USA E2579-E2586 (Sep. 4, 2012); Jinek et al., "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity" 337 Science 816-821 (Aug. 17, 2012); Carroll, "A CRISPR Approach to Gene Targeting" 20(9) Molecular Therapy 1658-1660 (Sep. 2012); U.S. Appl. No. 61/652,086, filed May 25, 2012; Al-Attar et al., Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs): The Hallmark of an Ingenious Antiviral Defense Mechanism in Prokaryotes, Biol Chem. (2011) vol. 392, Issue 4, pp. 277-289; Hale et al., Essential Features and Rational Design of CRISPR RNAs That Function With the Cas RAMP Module Complex to Cleave RNAs, Molecular Cell, (2012) vol. 45, Issue 3, 292-302.

The variant proteins described herein can be used in place of the SpCas9 proteins described in the foregoing references with guide RNAs that target sequences that have PAM sequences according to Tables 1, 2, or 3.

In addition, the variants described herein can be used in fusion proteins in place of the wild-type Cas9 or other Cas9 mutations (such as the dCas9 or Cas9 nickase described above) as known in the art, e.g., a fusion protein with a heterologous functional domains as described in WO 2014/124284. For example, the variants, preferably comprising one or more nuclease-reducing or killing mutation, can be fused on the N or C terminus of the Cas9 to a transcriptional activation domain or other heterologous functional domains (e.g., transcriptional repressors (e.g., KRAB, ERD, SID, and others, e.g., amino acids 473–530 of the ets2 repressor factor (ERF) repressor domain (ERD), amino acids 1–97 of the KRAB domain of KOX1, or amino acids 1–36 of the Mad mSIN3 interaction domain (SID); see Beerli et al., PNAS USA 95:14628-14633 (1998)) or silencers such as Heterochromatin Protein 1 (HP1, also known as swi6), e.g., HP1α or HP1β; proteins or peptides that could recruit long non-coding RNAs (lncRNAs) fused to a fixed RNA binding sequence such as those bound by the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein; enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or TET proteins); enzymes that modify histone subunits (e.g., histone acetyltransferases (HAT), histone deacetylases (HDAC), histone methyltransferases (e.g., for methylation of lysine or arginine residues) or histone demethylases (e.g., for demethylation of lysine or arginine residues)).

In some embodiments, the heterologous functional domain is a base editor, e.g., a deaminase that modifies cytosine DNA bases, e.g., a cytidine deaminase from the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like (APOBEC) family of deaminases, including APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D/E, APOBEC3F, APOBEC3G, APOBEC3H, and APOBEC4 (see, e.g., Yang et al., J Genet Genomics. 2017 Sep 20;44(9):423-437); activation-induced cytidine deaminase (AID), e.g., activation induced cytidine deaminase (AICDA); cytosine deaminase 1 (CDA1) and CDA2; and cytosine deaminase acting on tRNA (CDAT). The following table provides exemplary sequences; other sequences can also be used.

	GenBank Accession Nos.	
Deaminase	Nucleic Acid	Amino Acid
hAID/AICDA	NM_020661.3 isoform 1 NM_020661.3 isoform 2	NP_065712.1 variant 1 NP_065712.1 variant 2
APOBEC1	NM_001644.4 isoform a NM_005889.3 isoform b	NP_001635.2 variant 1 NP_005880.2 variant 3
APOBEC2	NM_006789.3	NP_006780.1
APOBEC3A	NM_145699.3 isoform a NM_001270406.1 isoform b	NP_663745.1 variant 1 NP_001257335.1 variant 2
APOBEC3B	NM_004900.4 isoform a NM_001270411.1 isoform b	NP_004891.4 variant 1 NP_001257340.1 variant 2
APOBEC3C	NM_014508.2	NP_055323.2
APOBEC3D/E	NM_152426.3	NP_689639.2
APOBEC3F	NM_145298.5 isoform a NM_001006666.1 isoform b	NP_660341.2 variant 1 NP_001006667.1 variant 2
APOBEC3G	NM_021822.3 (isoform a)	NP_068594.1 (variant 1)
APOBEC3H	NM_001166003.2	NP_001159475.2 (variant SV-200)
APOBEC4	NM_203454.2	NP_982279.1
CDA1*	NM_127515.4	NP_179547.1

* from *Saccharomyces cerevisiae* S288C

In some embodiments, the heterologous functional domain is a deaminase that modifies adenosine DNA bases, e.g., the deaminase is an adenosine deaminase 1 (ADA1), ADA2; adenosine deaminase acting on RNA 1 (ADAR1), ADAR2, ADAR3 (see, e.g., Savva et al., Genome Biol. 2012 Dec 28;13(12):252); adenosine deaminase acting on tRNA 1 (ADAT1), ADAT2, ADAT3 (see Keegan et al., RNA. 2017 Sep;23(9):1317-1328 and Schaub and Keller, Biochimie. 2002 Aug;84(8):791-803); and naturally occurring or engineered tRNA-specific adenosine deaminase (TadA) (see, e.g., Gaudelli et al., Nature. 2017 Nov 23;551(7681):464-471) (NP_417054.2 (Escherichia coli str. K-12 substr. MG1655); See, e.g., Wolf et al., EMBO J. 2002 Jul 15;21(14):3841-51). The following table provides exemplary sequences; other sequences can also be used.

	GenBank Accession Nos.	
Deaminase	Nucleic Acid	Amino Acid
ADA (ADA1)	NM_000022.3 variant 1	NP_000013.2 isoform 1
ADA2	NM_001282225.1	NP_001269154.1
ADAR	NM_001111.4	NP_001102.2
ADAR2 (ADARB1)	NM_001112.3 variant 1	NP_001103.1 isoform 1
ADAR3 (ADARB2)	NM_018702.3	NP_061172.1
ADAT1	NM_012091.4 variant 1	NP_036223.2 isoform 1
ADAT2	NM_182503.2 variant 1	NP_872309.2 isoform 1
ADAT3	NM_138422.3 variant 1	NP_612431.2 isoform 1

In some embodiments, the heterologous functional domain is an enzyme, domain, or peptide that inhibits or enhances endogenous DNA repair or base excision repair (BER) pathways, e.g., thymine DNA glycosylase (TDG; GenBank Acc Nos. NM_003211.4 (nucleic acid) and NP_003202.3 (protein)) or uracil DNA glycosylase (UDG, also known as uracil N-glycosylase, or UNG; GenBank Acc Nos. NM_003362.3 (nucleic acid) and NP_003353.1 (protein)) or uracil DNA glycosylase inhibitor (UGI) that inhibits UNG mediated excision of uracil to initiate BER (see, e.g., Mol et al., Cell 82, 701–708 (1995); Komor et al., Nature. 2016 May

19;533(7603)); or DNA end-binding proteins such as Gam, which is a protein from the bacteriophage Mu that binds free DNA ends, inhibiting DNA repair enzymes and leading to more precise editing (less unintended base edits). See, e.g., Komor et al., Sci Adv. 2017 Aug 30;3(8):eaao4774.

See, e.g., Komor et al., Nature. 2016 May 19;533(7603):420-4; Nishida et al., Science. 2016 Sep 16;353(6305). pii: aaf8729; Rees et al., Nat Commun. 2017 Jun 6;8:15790; or Kim et al., Nat Biotechnol. 2017 Apr;35(4):371-376) as are known in the art can also be used.

A number of sequences for domains that catalyze hydroxylation of methylated cytosines in DNA. Exemplary proteins include the Ten-Eleven-Translocation (TET)1-3 family, enzymes that converts 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC) in DNA.

Sequences for human TET1-3 are known in the art and are shown in the following table:

	GenBank Accession Nos.	
Gene	Amino Acid	Nucleic Acid
TET1	NP_085128.2	NM_030625.2
TET2*	NP_001120680.1 (var 1) NP_060098.3 (var 2)	NM_001127208.2 NM_017628.4
TET3	NP_659430.1	NM_144993.1

* Variant (1) represents the longer transcript and encodes the longer isoform (a). Variant (2) differs in the 5' UTR and in the 3' UTR and coding sequence compared to variant 1. The resulting isoform (b) is shorter and has a distinct C-terminus compared to isoform a.

In some embodiments, all or part of the full-length sequence of the catalytic domain can be included, e.g., a catalytic module comprising the cysteine-rich extension and the 2OGFeDO domain encoded by 7 highly conserved exons, e.g., the Tet1 catalytic domain comprising amino acids 1580-2052, Tet2 comprising amino acids 1290-1905 and Tet3 comprising amino acids 966-1678. See, e.g., Fig. 1 of Iyer et al., Cell Cycle. 2009 Jun 1;8(11):1698-710. Epub 2009 Jun 27, for an alignment illustrating the key catalytic residues in all three Tet proteins, and the supplementary materials thereof for full length sequences (see, e.g., seq 2c); in some embodiments, the sequence includes amino acids 1418-2136 of Tet1 or the corresponding region in Tet2/3.

Other catalytic modules can be from the proteins identified in Iyer et al., 2009.

In some embodiments, the heterologous functional domain is a biological tether, and comprises all or part of (e.g., DNA binding domain from) the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein. These proteins can be used to recruit RNA molecules containing a specific stem-loop structure to a locale specified by the dCas9 gRNA targeting sequences. For example, a dCas9 variant fused to MS2 coat protein, endoribonuclease Csy4, or lambda N can be used to recruit a long non-coding RNA (lncRNA) such as XIST or HOTAIR; see, e.g., Keryer-Bibens et al., Biol. Cell 100:125–138 (2008), that is linked to the Csy4, MS2 or lambda N binding sequence. Alternatively, the Csy4, MS2 or lambda N protein binding sequence can be linked to another protein, e.g., as described in Keryer-Bibens et al., supra, and the protein can be targeted to the dCas9 variant binding site using the methods and compositions described herein. In some embodiments, the Csy4 is catalytically inactive. In some embodiments, the Cas9 variant, preferably a dCas9 variant, is fused to FokI as described in WO 2014/204578.

In some embodiments, the fusion proteins include a linker between the dCas9 variant and the heterologous functional domains. Linkers that can be used in these fusion proteins (or between fusion proteins in a concatenated structure) can include any sequence that does not interfere with the function of the fusion proteins. In preferred embodiments, the linkers are short, e.g., 2-20 amino acids, and are typically flexible (i.e., comprising amino acids with a high degree of freedom such as glycine, alanine, and serine). In some embodiments, the linker comprises one or more units consisting of GGGS (SEQ ID NO:2) or GGGGS (SEQ ID NO:3), e.g., two, three, four, or more repeats of the GGGS (SEQ ID NO:2) or GGGGS (SEQ ID NO:3) unit. Other linker sequences can also be used.

Delivery and Expression Systems

To use the Cas9 variants described herein, it may be desirable to express them from a nucleic acid that encodes them. This can be performed in a variety of ways. For example, the nucleic acid encoding the Cas9 variant can be cloned into an intermediate vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding the Cas9 variant for production of the Cas9 variant. The nucleic acid encoding the Cas9 variant can also be cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

To obtain expression, a sequence encoding a Cas9 variant is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., Molecular Cloning, A Laboratory Manual (3d ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered protein are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., 1983, Gene 22:229-235). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

The promoter used to direct expression of a nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of fusion proteins. In contrast, when the Cas9 variant is to be administered in vivo for gene regulation, either a constitutive or an inducible promoter can be used, depending on the particular use of the Cas9 variant. In addition, a preferred promoter for administration of the Cas9 variant can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tetracycline-regulated systems and the RU-486 system (see, e.g., Gossen & Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547; Oligino et al., 1998, Gene Ther., 5:491-496; Wang et al., 1997, Gene Ther., 4:432-441; Neering et al., 1996, Blood, 88:1147-55; and Rendahl et al., 1998, Nat. Biotechnol., 16:757-761).

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the Cas9 variant, and any signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the Cas9 variant, e.g., expression in plants, animals, bacteria, fungus, protozoa, etc. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available tag-fusion expression systems such as GST and LacZ.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

The vectors for expressing the Cas9 variants can include RNA Pol III promoters to drive expression of the guide RNAs, e.g., the H1, U6 or 7SK promoters. These human promoters allow for expression of Cas9 variants in mammalian cells following plasmid transfection.

Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with the gRNA encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (see, e.g., Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (see, e.g., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the Cas9 variant.

Alternatively, the methods can include delivering the Cas9 variant protein and guide RNA together, e.g., as a complex. For example, the Cas9 variant and gRNA can be can be overexpressed in a host cell and purified, then complexed with the guide RNA (e.g., in a test tube) to form a ribonucleoprotein (RNP), and delivered to cells. In some embodiments, the variant Cas9 can be expressed in and purified from bacteria through the use of bacterial Cas9 expression plasmids. For example, His-tagged variant Cas9 proteins can be expressed in bacterial cells and then purified using nickel affinity chromatography. The use of RNPs circumvents the necessity of delivering plasmid DNAs encoding the nuclease or the guide, or encoding the nuclease as an mRNA. RNP delivery may also improve specificity, presumably because the half-life of the RNP is shorter and there's no persistent expression of the nuclease and guide (as you'd get from a plasmid). The RNPs can be delivered to the cells in vivo or in vitro, e.g., using lipid-mediated transfection or electroporation. See, e.g., Liang et al. "Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection." Journal of biotechnology 208 (2015): 44-53; Zuris, John A., et al. "Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo." Nature biotechnology 33.1 (2015): 73-80; Kim et al. "Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins." Genome research 24.6 (2014): 1012-1019.

The present invention includes the vectors and cells comprising the vectors.

EXAMPLES

The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

Methods

The following materials and methods were used in Example 1.

Plasmids and oligonucleotides

Sequences of oligonucleotides used to amplify endogenous human gene target sites for T7E1 assays are found in **Table 4**.

TABLE 4

Primers used for T7E1 experiments		SEQ ID NO:
sequence	description	
GGAGCAGCTGGTCAGAGGGG	forward primer targeted to *EMX1* in U2OS human cells	4
CCATAGGGAAGGGGGACACTGG	reverse primer targeted to *EMX1* in U2OS human cells	5
GGGCCGGGAAAGAGTTGCTG	forward primer targeted to *FANCF* in U2OS human cells	6
GCCCTACATCTGCTCTCCCTCC	reverse primer targeted to *FANCF* in U2OS human cells	7
CCAGCACAACTTACTCGCACTTGAC	forward primer targeted to *RUNX1* in U2OS human cells	8
CATCACCAACCCACAGCCAAGG	reverse primer targeted to *RUNX1* in U2OS human cells	9
GATGAGGGCTCCAGATGGCAC	forward primer targeted to *VEGFA* in U2OS human cells	10
GAGGAGGGAGCAGGAAAGTGAGG	reverse primer targeted to *VEGFA* in U2OS human cells	11

Bacterial Cas9/sgRNA expression plasmids were constructed as previously described (Kleinstiver et al., Nature 2015) with two T7 promoters to separately express Cas9 and the sgRNA. Bacterial expression plasmids containing variable amino acids at positions D1135, S1136, G1218, E1219, R1335, and T1337 were generated by cloning oligonucleotides encoding randomized codons at these positions into the parental SpCas9 bacterial expression vectors (Figure 1).

For expression in human cells, point mutations in SpCas9 were generated by isothermal assembly into a pCMV-T7-hSpCas9-NLS-3xFLAG vector (JDS246; sequences found here at addgene.org/43861/sequences/).

Plasmids for U6 expression of sgRNAs (into which desired spacer oligos can be cloned) were generated by cloning appropriate annealed oligos into BsmBI digested BPK1520.

Bacterial-based positive selection assay for evolving SpCas9 variants

Competent *E.coli* BW25141(λDE3)[23] containing a positive selection plasmid (with embedded target site) were transformed with Cas9/sgRNA-encoding plasmids.

Following a 60 minute recovery in SOB media, transformations were plated on LB plates containing either chloramphenicol (non-selective) or chloramphenicol + 10 mM arabinose (selective).

To select for SpCas9 variants that can cleave novel PAMs, plasmids encoding randomized D1135X/S1136X/G1218X/E1219X/R1335X/T1337X SpCas9 librarieswere electroporated into *E.coli* BW25141(λDE3) cells that already harbored a positive selection plasmid that encodes a target site with a PAM of interest. Surviving colonies were grown overnight, miniprepped to extract the SpCas9-XXXXXX expression plasmid, and retransformed individually into *E.coli* BW25141(λDE3) cells containing a positive selection with the previously described PAM sequence to re-test linkage of the survival phenotype to those plasmids and thereby eliminate false positive clones. Generally ~300 clones were re-screened in follow-up experiments. The SpCas9 expression plasmids of bona fide surviving colonies in the secondary screen were sequenced to identify the amino acids at positions D1135, S1136, G1218, E1219, R1335, and/or T1337 that led to the alteration in specificity (see Tables 1-3). Mutations observed in the sequenced clones were chosen for further assessment based on their frequency in surviving clones, and (in some cases) activities in a human cell-based EGFP disruption assay.

Human cell culture and transfection

U2OS cells and U2OS.EGFP cells harboring a single integrated copy of an EGFP-PEST reporter gene (Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460-465 (2012)) were cultured in Advanced DMEM medium (Life Technologies) with 10% FBS, penicillin/streptomycin, and 2 mM GlutaMAX (Life Technologies) at 37°C with 5% CO2. Cell line identities were validated by STR profiling (ATCC) and deep sequencing, and cells were tested bi-weekly for mycoplasma contamination. U2OS.EGFP culture medium was additionally supplemented with 400 μg/mL G418. Cells were co-transfected with 750 ng Cas9 plasmid and 250 ng sgRNA plasmid using the DN-100 program of a Lonza 4D-nucleofector following the manufacturer's instructions. Cas9 plasmid transfected together with an empty U6 promoter plasmid was used as a negative control for all human cell experiments.

Human cell EGFP disruption assay

EGFP disruption experiments were performed as previously described (Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31, 822-826 (2013); Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460-465 (2012)). Approximately 52 hours post-transfection, a Fortessa flow cytometer (BD Biosciences) was used to measure EGFP fluorescence in transfected U2OS.EGFP cells. Negative control transfections of Cas9 and empty U6 promoter plasmids were used to establish background EGFP loss at ~2.5% for all experiments.

T7E1 assay

T7E1 assays were performed as previously described[15]. For U2OS human cells, genomic DNA was extracted from transfected cells ~72 hours post-transfection using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter Genomics). Target loci from human cell genomic DNA were amplified using the primers listed in **Table 4**. Roughly 200 ng of purified PCR product was denatured, annealed, and digested with T7E1 (New England BioLabs). Mutagenesis frequencies were quantified using a Qiaxcel capillary electrophoresis instrument (QIagen), as previously described[15].

Example 1

To further expand the utility of SpCas9 by enabling targeting of currently inaccessible PAM sequences, we conceived of an alternate strategy to select for SpCas9 variants capable of recognizing novel PAM sequences. Having established previously that certain positions within the SpCas9 coding sequence are important for PAM recognition (Kleinstiver et al., Nature 2015), we conducted a focused mutagenesis approach where we randomized the amino acids at six positions to generate a library of SpCas9 variants with diverse codon usage within three separate regions of the PAM interacting domain: D1135/S1136, G1218/E1219, and R1335/T1337. To do so, we sequentially cloned in randomized oligonucleotide cassettes encoding NNS nucleotide triplets (where N is any nucleotide and S is G or C) at the codons of SpCas9 that contain encode these six amino acids (Figure 1A). The resulting library of SpCas9 variants was then screened in our bacterial positive selection assay as previously described (Kleinstiver et al., Nature 2015), against target sites that encode various NGNN PAM sequences (Figure 1B). Briefly, bacteria can only survive selective conditions (plating on 10 mM arabinose, which induces transcription of the ccdB toxic gene) if SpCas9 can recognize the target site (PAM and spacer sequence) encoded in the positive selection plasmid. Strong PAM recognition will lead to hydrolysis of the selection plasmid, preventing induction of ccdB expression and allowing bacterial growth. Thus, while screening SpCas9 libraries, colonies that grow on media containing 10 mM arabinose are expected to encode an SpCas9 PAM variant that can target alternate PAMs (Figure 1B).

We first screened the randomized D1135X/S1136X/G1218X/E1219X/R1335X/T1337X SpCas9 library (SpCas9-XXXXXX, where X is any amino acid) on positive selection plasmids that encode target sites with NGTG, NGTT, NGTC, and NGTA PAMs. For each different PAM selection, 48 surviving colonies from the arabinose selection were picked and grown overnight in chloramphenicol containing media to recover the nuclease encoding plasmid. To reduce false positive rate in the primary screen, all putative PAM variant plasmids were subsequently re-screened against positive selection plasmids encoding the target site and PAM against which they were originally screened (data not shown). A subset of bona fide variants with at least 50% survival in this re-screening assay were sequenced to identify the amino acids at residues 1135, 1136, 1218, 1219, 1335, 1337 (**Table 1**), and then these variants were screened more broadly against NGTG, NGTT, NGTC, and NGTA PAMs to assess activity against NGTN sites (**Figures 2a-b** and **Table 1**). **Note:** subsequent to this point, in bacterial assays SpCas9 variants will be described by their variant number (vNGTN-#) or in human assays by their ‘amino acid name’, where the amino acid name will be in the form SpCas9-XXXXXX with the six Xs representing the amino acids identities at positions 1135, 1136, 1218, 1219, 1335, and 1337 (found in **Tables 1-3**). Re-screening identified a few trends, where in some cases the variant had the highest activity on the NGTN PAM against which it was originally selected (ex. vNGTN-1, -3, -12, -27, -28, etc.), that some variants could target a combination of NGTN PAMs (ex. vNGTN-15, -31, -35, etc., that can target NGTC and NGTA), and some variants can target all NGTN PAMs (ex. vNGTN-9, -10, -30, etc.) (**Figures 2A-B** and **Table 1**). Based on these results, novel variants were rationally designed based on frequently occurring amino acids at positions 1135, 1136, 1218, 1219, 1335, and 1337 in the clones that performed well in the initial screens. These rationally designed NGTN variants were assessed in the bacterial screen against NGTG, NGTT, NGTC, and NGTA PAMs (**Figures 2C-F**), and in some cases also screened against NGAN PAMs (NGAG, NGAT, NGAC, NGAA; **Figures 2C, 2E-F**). Again a number of interesting variants were identified with properties consistent with the preferences above, but notably some additional variants were identified that could impart a preference on the 4th position of the PAM (ex., vNGTG-37 that could target NGT<u>G</u> or NGA<u>G</u> PAMs or vNGTG-18 and -41 on NGT<u>C</u> and NGA<u>C</u> PAMs, etc.), additional variants that can target all NGT<u>N</u> PAMs (ex. vNGTN-40, -46, -48, etc.), and variants that can target all or nearly all NGT<u>N</u> or NGA<u>N</u> PAMs (ex. vNGTN-7, -44, -59, etc.; **Figures 2C-F** and **Table 1**).

Having identified several variants that can target NGTN PAM sites in bacteria, we sought to determine whether these preferences would translate to bona fide activity in human cells. In in initial screen of twelve different NGTN PAM variants in our human U2OS *EGFP*-disruption assay, we identified variants that could robustly target NGT<u>T</u> and NGT<u>G</u> PAMs (ex. SpCas9-GRKIQK, - VAKLLR, -VRKLLR, etc.), and some that could modify all NGT<u>N</u> PAM sites (ex. SpCas9-LRSVQL, –IRAVQL, etc.) (**Figure 3A**). Further screening of a subset of these variants and additional rationally designed variants in the human cell *EGFP*-disruption assay identified SpCas9-LRSVQL, -LRKIQK, -LRSVQK, and others as promising variants that can target NGT<u>N</u> PAM sequences (**Figure 3B**). To more stringently characterize the activity of SpCas9-LRSVQL on NGT<u>N</u> PAM sequences in human cells, we examined the activity of this nuclease variant across 32 different endogenous sites across the *EMX1*, *FANCF*, and *RUNX1* genes in human U2OS cells. This analysis revealed robust activity of SpCas9-LRSVQL on various endogenous sites bearing NGTG, NGTA, NGTC, and NGTT PAMs (**Figure 3C**, demonstrating that our selected and rationally designed PAM variants can function efficiently across numerous loci not previously targetable with published SpCas9 variants.

We have previously described an SpCas9 variant that can effectively target NGCG PAM sites (Kleinstiver et al., *Nature*, 2015), called SpCas9-VRER (that encodes D1135V/G1218R/R1335E/T1337R substitutions). While this variant enables targeting of previously inaccessible sites, it is restricted to activity on sites with an extended NGCG PAM. To expand the utility of SpCas9 PAM variants by potentially targeting all NGC<u>N</u> PAMs to now include NGC<u>T</u>, NGC<u>C</u>, and NGC<u>A</u>, we performed similar selections to those described above, but screened the SpCas9-XXXXXX library against positive selection plasmids harboring a target site with either an NGCG, NGCT, NGCC, or NGCA PAM (**Figures 4A-B** and **Table 2**). Much like we observed with the NGTN selections, re-screening of NGCN variants identified cases where the variant had the highest activity on the NGCN PAM against which it was originally selected (ex. vNGCN-3, -8, -9, -17, etc.), that some variants could target a combination of NGCN PAMs (ex. vNGCN-10, etc., that can target NGCT, NGCC and NGCA), and some variants can target all NGCN PAMs (ex. vNGCN-1, -2, -5, -18, -26 etc.) (**Figures 4A-B** and **Table 2**). Various rationally generated NGCN variants were cloned based on observations of amino acid enrichment in SpCas9-XXXXXX selected clones, and tested in bacteria for activity against NGCN PAMs (**Figures 4A-B**,**Table 2**, and data not shown).

Next, we examined the activities of various NGCN selected SpCas9 PAM variants in our U2OS *EGFP*-disruption assay to determine whether their re-targeted PAM preferences and nuclease activities could be recapitulated in human cells (**Figures 5A-D**). We observed activity of numerous variants against NGCA PAMs (SpCas9-MQKSER, -LWRVVA, -LWLETR, etc.; **Figure 5A**), NGCC PAMs (SpCas9-MQKSER, -LSRWQR, -ICCCER, etc.; **Figure 5A**), NGCT PAMs (SpCas9-MQKSER; **Figure 5C**), or NGCC and NGCT PAMs (SpCas9-MQKSER, -VRKSER, -ICKSER, etc.; **Figure 5C**). Further testing of the SpCas9-MQKSER, -VRKSER, -ICKSER variants against 15 total NGCA, NGCC, NGCT, and NGCG sites revealed robust activity of each variant against all classes of NGCN PAMs (**Figure 5E**). In some cases, these variants can outperform the published SpCas9-VRER (e.g., as shown in **Figures 5B-C**), though this was generally on PAMs that were previously shown to be ineffectively targeted by SpCas9-VRER. Collectively, these new variants expand SpCas9 targeting to NGCT, NGCC, and NGCA instead of the formerly accessible NGCG, with SpCas9-MQKSER and other variants having robust activity on all NGCN PAMs.

Additionally, we have also previously described SpCas9 variant that can effectively target NGAN PAM sites, called SpCas9-VQR (D1135V/R1335Q/T1337R; Kleinstiver et al., *Nature*, 2015), and SpCas9-VRQR (D1135V/G1218R/R1335Q/T1337R; Kleinstiver and Pattanayak et al., *Nature*, 2016). However, these variants have a preference for subclasses of NGAN PAMs in the order of NGAG > NGAA = NGAT > NGAC, i.e., they have suboptimal activity against NGA<u>C</u> PAM sites. To potentially improve SpCas9 targeting of NGAN PAMs, we performed selections with the SpCas9-XXXXXX library as described above on positive selection plasmids encoding NGAG, NGAT, NGAC, and NGAA PAMs (**Figures 6A-B** and **Table 3**). Re-screening of NGA<u>N</u> variants revealed clones that had the highest activity on the NGA<u>N</u> PAM against which it was originally selected (ex. vNGAN-1, -2, -17, -26 through -30, etc. on NGA<u>G</u>, vNGAN-32 on NGA<u>T</u>, vNGAN-4, -5, -40, -41, etc. on NGA<u>C</u>, etc.), that some variants could target a combination of NGA<u>N</u> PAMs (ex. vNGAN-20, -21, etc., that can target NGA<u>T</u> and NGA<u>C</u>, or vNGAN-22 that can target NGA<u>G</u> and NGA<u>C</u>), and some variants can target all NGA<u>N</u> PAMs (ex. vNGAN-3, -13, -25, -31 etc.) (**Figures 6A-B** and **Table 3**).

Because numerous SpCas9-XXXXXX variants revealed strong NGAC PAM targeting in the bacterial screen, many variants were sub-cloned into our human expression vector to examine activity in our human cell U2OS *EGFP*-disruption assay. An initial screen of a subset of variants against single NGAA, NGAC, NGAT and NGAG PAM sites in *EGFP* revealed that certain variants could potentially outperform SpCas9-VQR at sites harboring NGAC PAMs (**Figure 7A**). More extensive testing of variants from **Figure 7A** and additional selected variants revealed that multiple SpCas9 variants had improved activity relative to SpCas9-VRQR on some or all four of the NGAC PAM sites examined in the *EGFP* disruption assay (**Figure 7B**), including SpCas9-LRSVRS, -MRARKE, -SRQMRG, and others.

We then compared the activity of our SpCas9 variants to a recently described SpCas9 PAM variant called xCas9 that has a reported relaxed NG PAM preference (Hu et al., Nature volume 556, pages 57–63 (05 April 2018)). Consistent with our previous results, we observed robust nuclease targeting (between 15-50% as assessed by T7E1 assay) of sites with NGA PAMs with the VRQR variant (also known as VSREQR), of sites with NGCG PAMs with the VRER (also known as VSREER) and MQKSER variants, and of sites with NGT PAMs with the LRSVQL variant (**Figure 8**). However, with the xCas9 variant, no targeting of sites with NGA, NGCG, or NGT PAMs was observed at greater than 10% efficiency; furthermore, we observed that xCas9 was on average about 2-fold less effective at targeting sites with NGG PAMs as compared to wild-type SpCas9 (**Figure 8**). These results demonstrate that our

SpCas9 PAM variants are more effective nucleases against a variety of PAMs when compared to xCas9.

Example 2.

The ability to perform precise single base editing events has recently been demonstrated using engineered SpCas9 base editor (BE) constructs (see, e.g., Komor et al., Nature. 2016 May 19;533(7603):420-4; Nishida et al., Science. 2016 Sep 16;353(6305); Kim et al., Nat Biotechnol. 2017 Apr;35(4):371-376; Komor et al., Sci Adv. 2017 Aug 30;3(8):eaao4774; and Gaudelli et al., Nature. 2017 Nov 23;551(7681):464-471), which exploit the formation of SpCas9-gRNA formed R-loops that cause ssDNA accessibility of the non-target DNA strand. The fusion of heterologous cytidine or adenine deaminase enzymatic domains to SpCas9 can therefore act on the exposed ssDNA strand, leading to the efficient introduction of C to T changes (so-called cytosine base editors, or CBEs), or A to G (so-called adenosine base editors, or ABEs), respectively. Because cellular base-excision repair (BER) employs uracil DNA glycosylase (UDG; also known as uracil N-glycosylase, or UNG) to excise uracil bases, this endogenous process can effectively reverse edits generated by cytidine BEs because the deamination of cytidine leads to a uracil intermediate. Therefore, to improve the efficiency of cytidine BEs, heterologous effector domains such as uracil glycosylase inhibitor (UGI) can also be fused to SpCas9 to inhibit UDG, subverting the initiation of BER and increasing the effectiveness of cytidine BEs.

We therefore sought to determine whether the expanded targeting range of our SpCas9 PAM variants could improve the utility of base editors by enabling editing of previously inaccessible sites. To do so, we constructed BE3 (Komor et al., Nature. 2016 May 19;533(7603):420-4) PAM variants to generate CBEs capable of recognizing sites with NGA and NGT PAMs. We found that on sites with NGA PAMs the CBE-VRQR variant exhibited between 7.5% to 64.2% conversion of Cs to Ts in the editing window, whereas xCas9 exhibited 0%-19.9% C-to-T editing on the same sites (**Figure 9A**). Similarly, on sites with NGT PAMs the CBE-LRSVQL variant exhibited between 10.8% to 50.3% conversion of Cs to Ts, whereas CBE-xCas9 exhibited 0%-28.5% C-to-T editing on the same sites (**Figure 9A**). We also observed a marked decrease in C-to-T editing activity with CBE-xCas9 (26.7%-37.2%) compared to wild-type SpCas9 (52.5%-62.4%) on sites with NGG PAMs

(**Figure 9A**). These results demonstrate that the BE3 versions of VRQR and LRSVQL are effective CBEs on sites with NGA and NGT PAMs, respectively, at rates ~2-fold greater than with xCas9.

Next, we constructed ABE(7.10) (Gaudelli et al., Nature. 2017 Nov 23;551(7681):464-471) versions of our PAM variants to determine their effectiveness as ABEs that mediate A-to-G conversion in human cells. We observed strong A-to-G editing activity with ABE-VRQR (8.0%-77.3%) on sites with NGA PAMs, compared to 0%-12.5% editing observed with ABE-xCas9 (**Figure 9B**). Similarly, on sites with NGCG PAMs, the ABE-VRER (0%-75.9%) and ABE-MSQKER (5.4%-90.4%) variants once again outperformed ABE-xCas9 (0%-62.3%) for A-to-G editing (**Figure 9B**). We also observed decreased A-to-G editing with ABE-xCas9 (0%-16.9%) compared to wild-type SpCas9 (13.9%-50.4%) on sites with NGG PAMs (**Figure 9B**). Our results reveal that the ABE(7.10) version of VRQR is effective at mediating A-to-G editing on sites with NGA PAMs, and that the ABE(7.10) versions of VRER and MSQKER are effective on sites with NGCG PAMs.

References

1. Sander, J.D. & Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32, 347-355 (2014).
2. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262-1278 (2014).
3. Doudna, J.A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258096 (2014).
4. Barrangou, R. & May, A.P. Unraveling the potential of CRISPR-Cas9 for gene therapy. Expert Opin Biol Ther 15, 311-314 (2015).
5. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012).
6. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. & Doudna, J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62-67 (2014).
7. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31, 827-832 (2013).

8. Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol 33, 187-197 (2015).

9. Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci U S A (2013).

10. 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).

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

12. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013).

13. Horvath, P. et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol 190, 1401-1412 (2008).

14. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569-573 (2014).

15. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460-465 (2012).

16. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31, 822-826 (2013).

17. Chen, Z. & Zhao, H. A highly sensitive selection method for directed evolution of homing endonucleases. Nucleic Acids Res 33, e154 (2005).

18. Doyon, J.B., Pattanayak, V., Meyer, C.B. & Liu, D.R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. J Am Chem Soc 128, 2477-2484 (2006).

19. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31, 233-239 (2013).

20. Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013).

21. Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31, 227-229 (2013).

22. Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol 10, 726-737 (2013).

23. Kleinstiver, B.P., Fernandes, A.D., Gloor, G.B. & Edgell, D.R. A unified genetic, computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI. Nucleic Acids Res 38, 2411-2427 (2010).

24. Gagnon, J.A. et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS One 9, e98186 (2014).

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

Table 1: Selection results and activity in bacteria of variants against NGTN PAMs

		Approximate survival in bacterial assay against:					Amino acid substitutions in variant:					
Sample #	Originally selected against (NGTN)	NGTG	NGTT	NGTC	NGTA	Variant name	D1135	S1136	G1218	E1219	R1335	T1337
vNGTN-1	G	30%	0%	3%	3%	SpCas9-DKVHVR	D	K	V	H	V	R
vNGTN-2	G	15%	40%	40%	40%	SpCas9-NRMMRT	N	R	M	M	R	T
vNGTN-3	G	100%	40%	40%	10%	SpCas9-ASQLMR	A	S	Q	L	M	R
vNGTN-4	G	100%	40%	10%	40%	SpCas9-LRQYTR	L	R	Q	Y	T	R
vNGTN-5	G	100%	15%	10%	20%	SpCas9-GCACMR	G	C	A	C	M	R
vNGTN-6	G	100%	50%	30%	20%	SpCas9-SRSCMV	S	R	S	C	M	V
vNGTN-7	T	100%	n/a	n/a	n/a	SpCas9-LWKIQK	L	W	K	I	Q	K
vNGTN-8	T	50%	50%	20%	30%	SpCas9-YSAFCC	Y	S	A	F	C	C
vNGTN-9	T	100%	100%	70%	80%	SpCas9-IRAVQL	I	R	A	V	Q	L
vNGTN-10	T	100%	80%	70%	60%	SpCas9-SWRVVV	S	W	R	V	V	V
vNGTN-11	T	100%	80%	70%	60%	SpCas9-SWKVLK	S	W	K	V	L	K
vNGTN-12	T	3%	80%	3%	3%	SpCas9-LWSVGG	L	W	S	V	G	G
vNGTN-13	C	-	-	10%	20%	SpCas9-SRAAKW	S	R	A	A	K	W
vNGTN-14	C	-	-	10%	10%	SpCas9-RNGWRI	R	N	G	W	R	I
vNGTN-15	C	0%	3%	90%	90%	SpCas9-TAHFKV	T	A	H	F	K	V
vNGTN-16	C	-	-	80%	80%	SpCas9-MSGVKC	M	S	G	V	K	C
vNGTN-17	C	-	-	50%	50%	SpCas9-VMRCKL	V	M	R	C	K	L
vNGTN-18	C	-	-	75%	75%	SpCas9-LRSVRS	L	R	S	V	R	S
vNGTN-19	A	-	-	n/a	30%	SpCas9-RWNLRR	R	W	N	L	R	R
vNGTN-20	A	3%	3%	0%	0%	SpCas9-VRCVRC	V	R	C	V	R	C
vNGTN-21	A	-	-	20%	20%	SpCas9-GRTSRC	G	R	T	S	R	C
vNGTN-22	A	-	-	65%	65%	SpCas9-LKLCKR	L	K	L	C	K	R
vNGTN-23	A	-	-	70%	65%	SpCas9-AKLCRT	A	K	L	C	R	T
vNGTN-24	A	-	-	75%	100%	SpCas9-SKTLRP	S	K	T	L	R	P
vNGTN-25	G	50%	20%	40%	40%	SpCas9-SRRSQR	S	R	R	S	Q	R
vNGTN-26	G	50%	20%	40%	50%	SpCas9-DKVHVR	D	K	V	H	V	R
vNGTN-27	G	50%	-	-	-	SpCas9-GAKLLR	G	A	K	L	L	R
vNGTN-28	T	20%	100%	40%	40%	SpCas9-MWAFGC	M	W	A	F	G	C
vNGTN-29	T	-	35%	-	-	SpCas9-GWRVTW	G	W	R	V	T	W
vNGTN-30	T	85%	100%	100%	80%	SpCas9-MWVHLN	M	W	V	H	L	N
vNGTN-31	C	-	-	100%	85%	SpCas9-TWSMRG	T	W	S	M	R	G
vNGTN-32	C	-	-	80%	35%	SpCas9-ISGTKN	I	S	G	T	K	N
vNGTN-33	C	-	-	50%	45%	SpCas9-SRAAKW	S	R	A	A	K	W
vNGTN-34	A	75%	75%	50%	40%	SpCas9-KCAFCC	K	C	A	F	C	C

Table 1: Selection results and activity in bacteria of variants against NGTN PAMs

Sample #	Originally selected against (NGTN)	Approximate survival in bacterial assay against:				Variant name	Amino acid substitutions in variant:					
		NGTG	NGTT	NGTC	NGTA		D1135	S1136	G1218	E1219	R1335	T1337
vNGTN-35	A	-	-	100%	100%	SpCas9-KRRCKV	K	R	R	C	K	V
vNGTN-36	A	-	-	90%	100%	SpCas9-MWGGRC	M	W	G	G	R	C
vNGTN-37	created variant	80%	-	5%	3%	SpCas9-VSKLLR	V	S	K	L	L	R
vNGTN-38	created variant	90%	20%	10%	10%	SpCas9-VRKLLR	V	R	K	L	L	R
vNGTN-27*	G	75%	-	-	-	SpCas9-GAKLLR	G	A	K	L	L	R
vNGTN-39	created variant	50%	95%	5%	2%	SpCas9-VSAVQL	V	S	A	V	Q	L
vNGTN-40	created variant	90%	95%	95%	50%	SpCas9-VRAVQL	V	R	A	V	Q	L
vNGTN-9*	T	90%	95%	95%	90%	SpCas9-IRAVQL	I	R	A	V	Q	L
vNGTN-41	created variant	-	-	95%	90%	SpCas9-VSSVRS	V	S	S	V	R	S
vNGTN-42	created variant	-	-	95%	90%	SpCas9-VRSVRS	V	R	S	V	R	S
vNGTN-18*	C	-	-	100%	95%	SpCas9-LRSVRS	L	R	S	V	R	S
vNGTN-43	N/A	-	-	-	-	SpCas9-SRGERT	S	R	G	E	R	T
vNGTN-44	N/A	80%	35%	100%	90%	SpCas9-SRMHCK	S	R	M	H	C	K
vNGTN-45	created variant					SpCas9-GRKIQK	G	R	K	I	Q	K
vNGTN-46	created variant					SpCas9-GWKLLR	G	W	K	L	L	R
vNGTN-47	created variant					SpCas9-GWKOQK	G	W	K	Q	Q	K
vNGTN-48	created variant					SpCas9-VAKLLR	V	A	K	L	L	R
vNGTN-49	created variant					SpCas9-VAKIQK	V	A	K	I	Q	K
vNGTN-50	created variant					SpCas9-VAKILR	V	A	K	I	L	R
vNGTN-51	created variant					SpCas9-GRKILR	G	R	K	I	L	R
vNGTN-52	created variant	-	-	100%	90%	SpCas9-VRKLRS	V	R	K	L	R	S
vNGTN-38	created variant	100%	85%	60%	50%	SpCas9-VRKLLR	V	R	K	L	L	R
vNGTN-53	created variant	100%	100%	100%	100%	SpCas9-LRSVQL	L	R	S	V	Q	L
vNGTN-18	C	-	1%	100%	100%	SpCas9-LRSVRS	L	R	S	V	R	S
vNGTN-54	created variant	-	5%	100%	100%	SpCas9-IRAVRS	I	R	A	V	R	S
vNGTN-55	T	100%	100%	95%	95%	SpCas9-IRAVQL	I	R	A	V	Q	L
vNGTN-56	created variant	-	-	50%	50%	SpCas9-VRKLKR	V	R	K	L	K	R
vNGTN-38	created variant	100%	50%	25%	25%	SpCas9-VRKLLR	V	R	K	L	L	R
vNGTN-57	created variant	-	-	100%	100%	SpCas9-SRSVRS	S	R	S	V	R	S
vNGTN-18	C	-	-	95%	90%	SpCas9-LRSVRS	L	R	S	V	R	S
vNGTN-58	created variant	100%	85%	100%	100%	SpCas9-VRKIQK	V	R	K	I	Q	K
vNGTN-7*	T	100%	85%	100%	100%	SpCas9-LWKIQK	L	W	K	I	Q	K
vNGTN-59	created variant	100%	85%	100%	100%	SpCas9-VRMHCK	V	R	M	H	C	K
vNGTN-44*	N/A	100%	60%	100%	95%	SpCas9-SRMHCK	S	R	M	H	C	K
vNGTN-61	created variant	n/a	n/a	n/a	n/a	SpCas9-GRKLLR	G	R	K	L	L	R
vNGTN-62	created variant	n/a	n/a	n/a	n/a	SpCas9-LRKIQK	L	R	K	I	Q	K

Table 1: Selection results and activity in bacteria of variants against NGTN PAMs

		Approximate survival in bacterial assay against:					Amino acid substitutions in variant:					
Sample #	Originally selected against (NGTN)	NGTG	NGTT	NGTC	NGTA	Variant name	D1135	S1136	G1218	E1219	R1335	T1337
vNGTN-63	created variant	n/a	n/a	n/a	n/a	SpCas9-LRSVQK	L	R	S	V	Q	K
vNGTN-64	created variant	n/a	n/a	n/a	n/a	SpCas9-VRKIQK	V	R	K	I	Q	K
vNGTN-65	created variant	n/a	n/a	n/a	n/a	SpCas9-GRSVQL	G	R	S	V	Q	L
vNGTN-66	created variant	n/a	n/a	n/a	n/a	SpCas9-GRKIQL	G	R	K	I	Q	L

*** = that the variant has already been screened in other experiments**

n/a = survival was not assessed in that experiment on that PAM

Table 2: Selection results and activity in bacteria of variants against NGCN PAMs

		Approximate survival in bacterial assay against:					Amino acid substitutions in variant:					
Sample #	Originally selected against (NGCN)	NGCG	NGCT	NGCC	NGCA	Variant name	D1135	S1136	G1218	E1219	R1335	T1337
vNGCN-1	G	100%	100%	100%	100%	SpCas9-WMQAYG	W	M	Q	A	Y	G
vNGCN-2	G	100%	100%	100%	n/a	SpCas9-MQKSER	M	Q	K	S	E	R
vNGCN-3	G	100%	-	-	40%	SpCas9-YSVCER	Y	S	V	C	E	R
vNGCN-4	T	90%	85%	90%	95%	SpCas9-CWNWNS	C	W	N	W	N	S
vNGCN-5	T	100%	100%	100%	100%	SpCas9-LWRSEY	L	W	R	S	E	Y
vNGCN-6	T	-	95%	95%	100%	SpCas9-QSTWNK	Q	S	T	W	N	K
vNGCN-7	C	n/a	n/a	n/a	n/a	SpCas9-LFEWRA	L	F	E	W	R	A
vNGCN-8	C	-	-	100%	-	SpCas9-SQSWRS	S	Q	S	W	R	S
vNGCN-9	C	-	-	100%	-	SpCas9-LKAWRS	L	K	A	W	R	S
vNGCN-10	A	-	100%	100%	100%	SpCas9-LWGWQH	L	W	G	W	Q	H
vNGCN-11	A	-	15%	15%	95%	SpCas9-LSYWAK	L	S	Y	W	A	K
vNGCN-12	A	50%	10%	20%	95%	SpCas9-RQMYQG	R	Q	M	Y	Q	G
vNGCN-13	created variant	-	-	-	-	SpCas9-LWREER	L	W	R	E	E	R
vNGCN-14	created variant	100%	5%	10%	20%	SpCas9-VSSWER	V	S	S	W	E	R
vNGCN-15	created variant	100%	3%	5%	15%	SpCas9-VSAWER	V	S	A	W	E	R
vNGCN-16	created variant	-	-	-	-	SpCas9-DWREER	D	W	R	E	E	R
vNGCN-17	created variant	100%	-	-	-	SpCas9-VSGWER	V	S	G	W	E	R
vNGCN-18	G	100%	100%	100%	100%	SpCas9-MCSFER	M	C	S	F	E	R
vNGCN-19	G	100%	-	-	25%	SpCas9-VLMYER	V	L	M	Y	E	R
vNGCN-20	G	100%	n/a	n/a	n/a	SpCas9-QGANER	Q	G	A	N	E	R
vNGCN-21	G	100%	50%	15%	50%	SpCas9-GCACER	G	C	A	C	E	R
vNGCN-22	G	100%	-	-	5%	SpCas9-SRIAER	S	R	I	A	E	R

Table 2: Selection results and activity in bacteria of variants against NGCN PAMs

Sample #	Originally selected against (NGCN)	Approximate survival in bacterial assay against: NGCG	NGCT	NGCC	NGCA	Variant name	Amino acid substitutions in variant: D1135	S1136	G1218	E1219	R1335	T1337
vNGCN-23	G	100%	-	-	25%	SpCas9-SRRNER	S	R	R	N	E	R
vNGCN-10*	T	-	100%	90%	100%	SpCas9-LWGWQH	L	W	G	W	Q	H
vNGCN-24	T	-	5%	-	-	SpCas9-WMQAVV	W	M	Q	A	V	V
vNGCN-25	T	-	100%	-	75%	SpCas9-AYRWSK	A	Y	R	W	S	K
vNGCN-26	T	100%	100%	30%	65%	SpCas9-LWMREQ	L	W	M	R	E	Q
vNGCN-27	T	-	100%	5%	50%	SpCas9-LWRVVA	L	W	R	V	V	A
vNGCN-28	T	100%	100%	n/a	75%	SpCas9-HSSWVR	H	S	S	W	V	R
vNGCN-29	C	100%	100%	100%	85%	SpCas9-MWSEPT	M	W	S	E	P	T
vNGCN-30	C	100%	100%	50%	80%	SpCas9-GWSMQR	G	W	S	M	Q	R
vNGCN-31	C	-	n/a	75%	-	SpCas9-NKAWRV	N	K	A	W	R	V
vNGCN-32	C	75%	-	95%	50%	SpCas9-LCTYEY	L	C	T	Y	E	Y
vNGCN-33	C	80%	5%	50%	50%	SpCas9-GSNWCK	G	S	N	W	C	K
vNGCN-34	C	85%	50%	90%	100%	SpCas9-GSNYQS	G	S	N	Y	Q	S
vNGCN-35	A	n/a	50%	25%	90%	SpCas9-FMQWVN	F	M	Q	W	V	N
vNGCN-36	A	40%	50%	75%	100%	SpCas9-YCSWVG	Y	C	S	W	V	G
vNGCN-37	A	50%	-	25%	85%	SpCas9-LWKFEG	L	W	K	F	E	G
vNGCN-38	A	25%	35%	5%	100%	SpCas9-MCAWCG	M	C	A	W	C	G
vNGCN-39	A	50%	-	50%	50%	SpCas9-GKNWNR	G	K	N	W	N	R
vNGCN-2*	A	100%	25%	25%	75%	SpCas9-MQKSER	M	Q	K	S	E	R
vNGCN-40	created variant	n/a	n/a	n/a	n/a	SpCas9-VRREER	V	R	R	E	E	R
vNGCN-41	A	n/a	n/a	n/a	n/a	SpCas9-AARWCQ	A	A	R	W	C	Q
vNGCN-42	A	n/a	n/a	n/a	n/a	SpCas9-LWLETR	L	W	L	E	T	R
vNGCN-43	A	n/a	n/a	n/a	85%	SpCas9-FMQWVR	F	M	Q	W	V	R
vNGCN-44	A	n/a	n/a	n/a	75%	SpCas9-SSKWPA	S	S	K	W	P	A
vNGCN-45	C	n/a	n/a	50%	n/a	SpCas9-MWASEG	M	W	A	S	E	G
vNGCN-46	A	n/a	n/a	n/a	100%	SpCas9-LSRWQR	L	S	R	W	Q	R
vNGCN-47	G	90%	n/a	n/a	n/a	SpCas9-YAIYER	Y	A	I	Y	E	R
vNGCN-48	G	75%	n/a	n/a	n/a	SpCas9-ICCCER	I	C	C	C	E	R
vNGCN-49	G	95%	n/a	n/a	n/a	SpCas9-DWFYER	D	W	F	Y	E	R
vNGCN-50	G	80%	n/a	n/a	n/a	SpCas9-REATER	R	E	A	T	E	R
vNGCN-51	G	75%	n/a	n/a	n/a	SpCas9-GWAYER	G	W	A	Y	E	R
vNGCN-52	G	75%	n/a	n/a	n/a	SpCas9-YAIYER	Y	A	I	Y	E	R
vNGCN-53	G	85%	n/a	n/a	n/a	SpCas9-LSVSER	L	S	V	S	E	R
vNGCN-54	A	n/a	n/a	n/a	75%	SpCas9-VRAWCR	V	R	A	W	C	R
vNGCN-55	G	75%	n/a	n/a	n/a	SpCas9-KWREQR	K	W	R	E	Q	R
vNGCN-56	G	75%	n/a	n/a	n/a	SpCas9-ARGAER	A	R	G	A	E	R

Table 2: Selection results and activity in bacteria of variants against NGCN PAMs

Sample #	Originally selected against (NGCN)	Approximate survival in bacterial assay against: NGCG	NGCT	NGCC	NGCA	Variant name	Amino acid substitutions in variant: D1135	S1136	G1218	E1219	R1335	T1337
vNGCN-57	C	n/a	n/a	75%	n/a	SpCas9-HASWCK	H	A	S	W	C	K
vNGCN-58	G	100%	n/a	n/a	n/a	SpCas9-YVRSER	Y	V	R	S	E	R
vNGCN-59	G	80%	n/a	n/a	n/a	SpCas9-QRLAER	Q	R	L	A	E	R
vNGCN-60	A	n/a	n/a	n/a	n/a	SpCas9-AARWER	A	A	R	W	E	R
vNGCN-61	G	75%	n/a	n/a	n/a	SpCas9-LILSER	L	I	L	S	E	R
vNGCN-62	A	n/a	n/a	n/a	n/a	SpCas9-LWPSRG	L	W	P	S	R	G
vNGCN-63	A	n/a	n/a	n/a	n/a	SpCas9-LWTWIK	L	W	T	W	I	K
vNGCN-64	created variant	n/a	n/a	n/a	n/a	SpCas9-VRKSER	V	R	K	S	E	R
vNGCN-65	created variant	n/a	n/a	n/a	n/a	SpCas9-ICKSER	I	C	K	S	E	R
vNGCN-66	C/T	n/a	n/a	n/a	n/a	SpCas9-MQSVQL	M	Q	S	V	Q	L
vNGCN-67	created variant	n/a	n/a	n/a	n/a	SpCas9-LRSVER	L	R	S	V	E	R
vNGCN-68	created variant	n/a	n/a	n/a	n/a	SpCas9-LSRWER	L	S	R	W	E	R

*** = that the variant has already been screened in other experiments**

n/a = survival was not assessed in that experiment on that PAM

Table 3: Selection results and activity in bacteria of variants against NGAN PAMs

Sample #	Originally selected against (NGAN)	Approximate survival in bacterial assay against: NGAG	NGAT	NGAC	NGAA	Variant name	Amino acid substitutions in variant: D1135	S1136	G1218	E1219	R1335	T1337
vNGAN-1	G	100%	10%	5%	1%	SpCas9-LRLSAR	L	R	L	S	A	R
vNGAN-2	G	100%	-	1%	-	SpCas9-ASEVTR	A	S	E	V	T	R
vNGAN-3	T	100%	100%	95%	90%	SpCas9-KWMMCG	K	W	M	M	C	G
vNGAN-4	C	-	-	100%	-	SpCas9-VRGAKE	V	R	G	A	K	E
vNGAN-5	C	-	-	100%	-	SpCas9-MRARKE	M	R	A	R	K	E
vNGAN-6	G	75%	-	-	-	SpCas9-AEEQQR	A	E	E	Q	Q	R
vNGAN-7	A	95%	5%	-	80%	SpCas9-TRGSFR	T	R	G	S	F	R
vNGAN-8	A	95%	10%	90%	90%	SpCas9-VRNYTK	V	R	N	Y	T	K
vNGAN-9	T	100%	100%	95%	95%	SpCas9-AWNFQV	A	W	N	F	Q	V
vNGAN-10	A	100%	35%	-	20%	SpCas9-WMRKVA	W	M	R	K	V	A
vNGAN-11	A	40%	100%	-	75%	SpCas9-CWTCLQ	C	W	T	C	L	Q
vNGAN-12	A	100%	100%	5%	75%	SpCas9-LWTTLN	L	W	T	T	L	N

Table 3: Selection results and activity in bacteria of variants against NGAN PAMs

Sample #	Originally selected against (NGAN)	Approximate survival in bacterial assay against:				Variant name	Amino acid substitutions in variant:					
		NGAG	NGAT	NGAC	NGAA		D1135	S1136	G1218	E1219	R1335	T1337
vNGAN-13	G	100%	95%	95%	95%	SpCas9-SRMHCK	S	R	M	H	C	K
vNGAN-14	T	100%	100%	95%	95%	SpCas9-CWCQCV	C	W	C	Q	C	V
vNGAN-15	T	100%	5%	-	10%	SpCas9-GCLCVR	G	C	L	C	V	R
vNGAN-16	C	100%	50%	-	-	SpCas9-GGCQLR	G	G	C	Q	L	R
vNGAN-17	G	100%	-	-	-	SpCas9-AEEQQR	A	E	E	Q	Q	R
vNGAN-18	G	90%	100%	10%	25%	SpCas9-QNNQVF	Q	N	N	Q	V	F
vNGAN-19	T	100%	100%	-	100%	SpCas9-GWEKVR	G	W	E	K	V	R
vNGAN-20	T	1%	100%	50%	-	SpCas9-NRAVNG	N	R	A	V	N	G
vNGAN-21	T	1%	100%	50%	-	SpCas9-NRAVNG	N	R	A	V	N	G
vNGAN-22	C	100%	1%	100%	-	SpCas9-SRQMRG	S	R	Q	M	R	G
vNGAN-23	C	-	-	-	-	SpCas9-RAQPNL	R	A	Q	P	N	L
vNGAN-24	A	50%	5%	-	100%	SpCas9-LRSYLH	L	R	S	Y	L	H
vNGAN-25	G	100%	95%	100%	90%	SpCas9-SRMHCK	S	R	M	H	C	K
vNGAN-26	G	100%	-	-	-	SpCas9-ACTSVR	A	C	T	S	V	R
vNGAN-27	G	100%	-	-	-	SpCas9-MVVHIR	M	V	V	H	I	R
vNGAN-28	G	100%	-	-	-	SpCas9-VRGNNR	V	R	G	N	N	R
vNGAN-29	G	100%	-	-	-	SpCas9-RGFCLR	R	G	F	C	L	R
vNGAN-30	G	100%	-	-	-	SpCas9-VQDAQR	V	Q	D	A	Q	R
vNGAN-31	T	100%	100%	95%	95%	SpCas9-GWRQSK	G	W	R	Q	S	K
vNGAN-32	T	5%	100%	-	-	SpCas9-AWLCLS	A	W	L	C	L	S
vNGAN-33	T	100%	100%	-	100%	SpCas9- KWARVV	K	W	A	R	V	V
vNGAN-34	T	80%	100%	20%	15%	SpCas9-LAAQTP	L	A	A	Q	T	P
vNGAN-35	T	95%	100%	10%	90%	SpCas9-GWNHLQ	G	W	N	H	L	Q
vNGAN-36	T	100%	100%	100%	5%	SpCas9-MWAARP	M	W	A	A	R	P
vNGAN-37	C	95%	100%	50%	30%	SpCas9-KWRCTG	K	W	R	C	T	G
vNGAN-38	C	50%	-	100%	-	SpCas9-LAKARP	L	A	K	A	R	P
vNGAN-39	C	100%	100%	100%	30%	SpCas9-SRMHCK	S	R	M	H	C	K
vNGAN-40	C	-	-	100%	-	SpCas9-VKMAKG	V	K	M	A	K	G
vNGAN-41	C	-	-	100%	-	SpCas9-QRKTRE	Q	R	K	T	R	E
vNGAN-42	C	-	-	50%	-	SpCas9-NTAVKQ	N	T	A	V	K	Q
vNGAN-43	A	100%	100%	50%	100%	SpCas9-LCRQQR	L	C	R	Q	Q	R
vNGAN-44	A	100%	90%	100%	100%	SpCas9-CWSHQR	C	W	S	H	Q	R
vNGAN-45	A	30%	90%	25%	100%	SpCas9-MWVHLN	M	W	V	H	L	N
vNGAN-46	A	100%	100%	25%	100%	SpCas9-SRTHTQ	S	R	T	H	T	Q
vNGAN-47	A	100%	50%	-	100%	SpCas9-LQKSMR	L	Q	K	S	M	R
vNGAN-48	A	100%	100%	-	90%	SpCas9-LWEVIR	L	W	E	V	I	R

Table 3: Selection results and activity in bacteria of variants against NGAN PAMs

Sample #	Originally selected against (NGAN)	Approximate survival in bacterial assay against: NGAG	NGAT	NGAC	NGAA	Variant name	Amino acid substitutions in variant: D1135	S1136	G1218	E1219	R1335	T1337
vNGTN-37	created variant	20%	-	-	-	SpCas9-VSKLLR	V	S	K	L	L	R
vNGTN-38	created variant	50%	-	-	-	SpCas9-VRKLLR	V	R	K	L	L	R
vNGTN-27*	NGTG	10%	-	-	-	SpCas9-GAKLLR	G	A	K	L	L	R
vNGTN-39	created variant	-	-	-	-	SpCas9-VSAVQL	V	S	A	V	Q	L
vNGTN-40	created variant	1%	-	-	-	SpCas9-VRAVQL	V	R	A	V	Q	L
vNGTN-9*	NGTT	5%	-	1%	-	SpCas9-IRAVQL	I	R	A	V	Q	L
vNGTN-41	created variant	1%	-	100%	-	SpCas9-VSSVRS	V	S	S	V	R	S
vNGTN-42	created variant	25%	-	100%	-	SpCas9-VRSVRS	V	R	S	V	R	S
vNGTN-18*	NGTC	25%	-	100%	-	SpCas9-LRSVRS	L	R	S	V	R	S
vNGTN-43	N/A	50%	-	-	1%	SpCas9-SRGERT	S	R	G	E	R	T
vNGTN-44	N/A	90%	80%	n/a	50%	SpCas9-SRMHCK	S	R	M	H	C	K
vNGTN-52	created variant	60%	-	75%	-	SpCas9-VRKLRS	V	R	K	L	R	S
vNGTN-38*	created variant	60%	-	-	-	SpCas9-VRKLLR	V	R	K	L	L	R
vNGTN-53	created variant	15%	5%	10%	1%	SpCas9-LRSVQL	L	R	S	V	Q	L
vNGTN-18*	NGTC	50%	5%	100%	-	SpCas9-LRSVRS	L	R	S	V	R	S
vNGTN-54	created variant	50%	5%	100%	-	SpCas9-IRAVRS	I	R	A	V	R	S
vNGTN-55	NGTT	5%	1%	1%	-	SpCas9-IRAVQL	I	R	A	V	Q	L
vNGTN-56	created variant	5%	-	35%	-	SpCas9-VRKLKR	V	R	K	L	K	R
vNGTN-38*	created variant	35%	-	-	-	SpCas9-VRKLLR	V	R	K	L	L	R
vNGTN-57	created variant	20%	-	100%	-	SpCas9-SRSVRS	S	R	S	V	R	S
vNGTN-18*	NGTC	25%	-	100%	-	SpCas9-LRSVRS	L	R	S	V	R	S
vNGTN-58	created variant	85%	5%	95%	25%	SpCas9-VRKIQK	V	R	K	I	Q	K
vNGTN-7*	NGTT	85%	100%	100%	95%	SpCas9-LWKIQK	L	W	K	I	Q	K
vNGTN-59	created variant	95%	100%	100%	90%	SpCas9-VRMHCK	V	R	M	H	C	K
vNGTN-44*	N/A	85%	90%	100%	75%	SpCas9-SRMHCK	S	R	M	H	C	K

*** = that the variant has already been screened in other experiments**

n/a = survival was not assessed in that experiment on that PAM

WHAT IS CLAIMED IS:

1. An isolated *Streptococcus pyogenes* Cas9 (SpCas9) protein, with mutations at three, four, five, or all six of the following positions: D1135, S1136, G1218, E1219, R1335, and/or T1337, wherein the mutations are not VSREER or VSREQR.

2. The isolated protein of claim 1, comprising a sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO:1.

3. The isolated protein of claim 1, comprising a set of mutations shown in Table A, 1, 2, or 3.

4. The isolated protein of claims 1-3, comprising the following sets of mutations: LRSVQL, LRKIQK, LRSVQK, LWKIQK, VRKIQK, LWKIQK, IRAVQL, VRKLRS, GRKIQK, SWRVVV, SWKVLK, TAHFKV, MSGVKC, LRSVRS, SKTLRP, MWVHLN, TWSMRG, KRRCKV, VRAVQL, VSSVRS, VRSVRS, SRMHCK, GWKLLR, GWKOQK, VAKLLR, VAKIQK, VAKILR, GRKILR, VRKLLR, IRAVQL, VRKIQK, or VRMHCK variant (e.g., for NGTN PAMs); MQKSER, VRKSER, ICKSER, LRSVER, LWLETR, LSRWER, MQSVQL, VRREER, ICCCER, LSRWQR, LWRVVA, WMQAYG, LWRSEY, SQSWRS, LKAWRS, LWGWQH, MCSFER, LWMREQ, LWRVVA, HSSWVR, MWSEPT, GSNYQS, FMQWVN, YCSWVG, MCAWCG, FMQWVR, or SSKWPA variant (e.g., for NGCN PAMs); LRSVRS, SRQMRG, MRARKE, SRMHCK, VRREQR, VRGEQR, LRLSAR, AWTEVTR, KWMMCG, VRGAKE, AWNFQV, LWTTLN, SRMHCK, CWCQCV, AEEQQR, GWEKVR, NRAVNG, LRSYLH, VRGNNR, VQDAQR, GWRQSK, AWLCLS, KWARVV, MWAARP, SRMHCK, VKMAKG, QRKTRE, LCRQQR, CWSHQR, SRTHTQ, LWEVIR, VSSVRS, VRSVRS, IRAVRS, SRSVRS, LWKIQK, or VRMHCK variant (e.g., for NGAN PAMs).

5. The isolated protein of claim 1, further comprising one or more mutations that decrease nuclease activity selected from the group consisting of mutations at D10, E762, D839, H983, or D986; and at H840 or N863.

6. The isolated protein of claim 5, wherein the mutations are:
   (i) D10A or D10N, and
   (ii) H840A, H840N, or H840Y.

7. The isolated protein of claim 1, further comprising one or more mutations that increase specificity selected from the group consisting of mutations at N497, R661, N692, M694, Q695, H698, K810, K848, Q926, K1003, and R0160, and optionally at K526 and/or R691.

8. The isolated protein of claim 7, comprising mutations N692A, Q695A, Q926A, H698A, N497A, R661A, M694A, K810A, K848A, K1003A, R0160A, Y450A/Q695A, L169A/Q695A, Q695A/Q926A, Q695A/D1135E, Q926A/D1135E, Y450A/D1135E, L169A/Y450A/Q695A, L169A/Q695A/Q926A, Y450A/Q695A/Q926A, R661A/Q695A/Q926A, N497A/Q695A/Q926A, Y450A/Q695A/D1135E, Y450A/Q926A/D1135E, Q695A/Q926A/D1135E, L169A/Y450A/Q695A/Q926A, L169A/R661A/Q695A/Q926A,Y450A/R661A/Q695A/Q926A, N497A/Q695A/Q926A/D1135E, R661A/Q695A/Q926A/D1135E, and Y450A/Q695A/Q926A/D1135E; N692A/M694A/Q695A/H698A, N692A/M694A/Q695A/H698A/Q926A; N692A/M694A/Q695A/Q926A; N692A/M694A/H698A/Q926A; N692A/Q695A/H698A/Q926A; M694A/Q695A/H698A/Q926A; N692A/Q695A/H698A; N692A/M694A/Q695A; N692A/H698A/Q926A; N692A/M694A/Q926A; N692A/M694A/H698A; M694A/Q695A/H698A; M694A/Q695A/Q926A; Q695A/H698A/Q926A; G582A/V583A/E584A/D585A/N588A/Q926A; G582A/V583A/E584A/D585A/N588A; T657A/G658A/W659A/R661A/Q926A; T657A/G658A/W659A/R661A; F491A/M495A/T496A/N497A/Q926A; F491A/M495A/T496A/N497A; K918A/V922A/R925A/Q926A; or 918A/V922A/R925A; K855A; K810A/K1003A/R1060A; or K848A/K1003A/R1060A, and optionally K526A and/or R691A.

9. A fusion protein comprising the isolated protein of claims 1-8, fused to a heterologous functional domain, with an optional intervening linker, wherein the linker does not interfere with activity of the fusion protein.

10. The fusion protein of claim 9, wherein the heterologous functional domain is a transcriptional activation domain.

11. The fusion protein of claim 9, wherein the transcriptional activation domain is from VP16, VP64, rTA, NF-κB p65, or the composite VPR (VP64-p65-rTA).

12. The fusion protein of claim 9, wherein the heterologous functional domain is a transcriptional silencer or transcriptional repression domain.

13. The fusion protein of claim 12, wherein the transcriptional repression domain is a Krueppel-associated box (KRAB) domain, ERF repressor domain (ERD), or mSin3A interaction domain (SID).

14. The fusion protein of claim 12, wherein the transcriptional silencer is Heterochromatin Protein 1 (HP1).

15. The fusion protein of claim 12, wherein the heterologous functional domain is an enzyme that modifies the methylation state of DNA.

16. The fusion protein of claim 15, wherein the enzyme that modifies the methylation state of DNA is a DNA methyltransferase (DNMT) or a TET protein.

17. The fusion protein of claim 16, wherein the TET protein is TET1.

18. The fusion protein of claim 9, wherein the heterologous functional domain is an enzyme that modifies a histone subunit.

19. The fusion protein of claim 18, wherein the enzyme that modifies a histone subunit is a histone acetyltransferase (HAT), histone deacetylase (HDAC), histone methyltransferase (HMT), or histone demethylase.

20. The fusion protein of claim 9, wherein the heterologous functional domain is a base editor.

21. The fusion protein of claim 20, wherein the base editor is (i) a cytidine deaminase domain, preferably selected from the group consisting of the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like (APOBEC) family of deaminases, optionally APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D/E,

APOBEC3F, APOBEC3G, APOBEC3H, or APOBEC4; activation-induced cytidine deaminase (AID), optionally activation induced cytidine deaminase (AICDA); cytosine deaminase 1 (CDA1) or CDA2; or cytosine deaminase acting on tRNA (CDAT), or (ii) an adenosine deaminase, preferably selected from the group consisting of adenosine deaminase 1 (ADA1), ADA2; adenosine deaminase acting on RNA 1 (ADAR1), ADAR2, ADAR3; adenosine deaminase acting on tRNA 1 (ADAT1), ADAT2, ADAT3; and naturally occurring or engineered tRNA-specific adenosine deaminase (TadA).

22. The fusion protein of claim 9, wherein the heterologous functional domain is a biological tether.

23. The fusion protein of claim 22, wherein the biological tether is MS2, Csy4 or lambda N protein.

24. The fusion protein of claim 9, wherein the heterologous functional domain is FokI.

25. An isolated nucleic acid encoding the protein of claims 1-24.

26. A vector comprising the isolated nucleic acid of claim 25.

27. The vector of claim 24, wherein the isolated nucleic acid of claim 25 is operably linked to one or more regulatory domains for expressing an isolated *Streptococcus pyogenes* Cas9 (SpCas9) protein, with mutations at one, two, three, four, five, or all six of the following positions: D1135, S1136, G1218, E1219, R1335, and/or T1337.

28. A host cell, preferably a mammalian host cell, comprising the nucleic acid of claim 25, and optionally expressing the protein of claims 1-24.

29. A method of altering the genome of a cell, the method comprising expressing in the cell, or contacting the cell with, the isolated protein or fusion protein of claim 1-24, and a guide RNA having a region complementary to a selected portion of the genome of the cell.

30. The method of claim 29, wherein the isolated protein or fusion protein comprises one or more of a nuclear localization sequence, cell penetrating peptide sequence, and/or affinity tag.

31. The method of claim 30, wherein the cell is a stem cell.

32. The method of claim 31, wherein the cell is an embryonic stem cell, mesenchymal stem cell, or induced pluripotent stem cell; is in a living animal; or is in an embryo.

33. A method of altering a double stranded DNA (dsDNA) molecule, the method comprising contacting the dsDNA molecule with the isolated protein or fusion protein of claims 1-24, and a guide RNA having a region complementary to a selected portion of the dsDNA molecule.

34. The method of claim 33, wherein the dsDNA molecule is *in vitro*.

35. The method of claim 33, wherein the fusion protein and RNA are in a ribonucleoprotein complex.

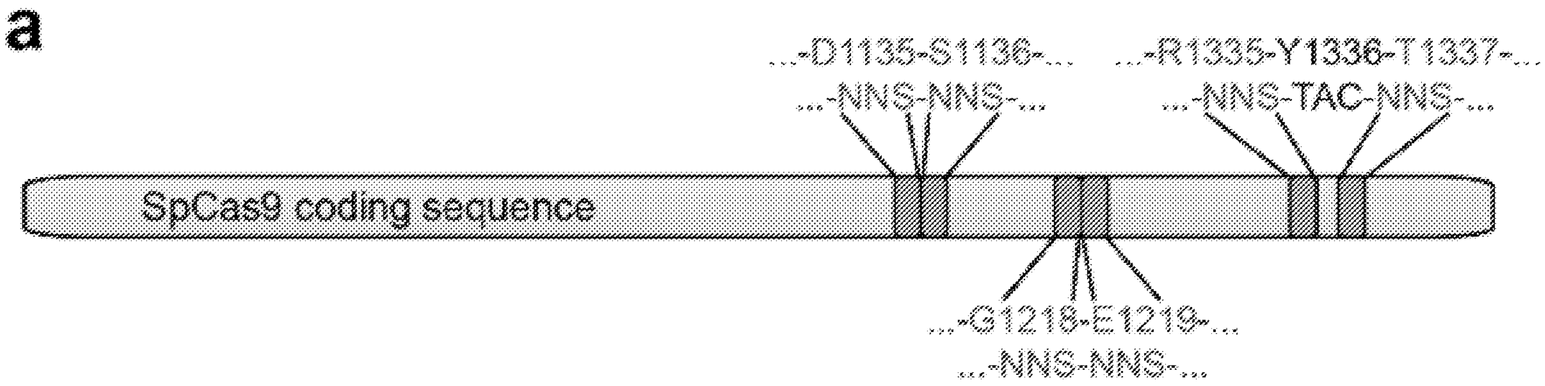

a
...-D1135-S1136-...
...-NNS-NNS-...
...-R1335-Y1336-T1337-...
...-NNS-TAC-NNS-...
SpCas9 coding sequence
...-G1218-E1219-...
...-NNS-NNS-...

FIG. 1A

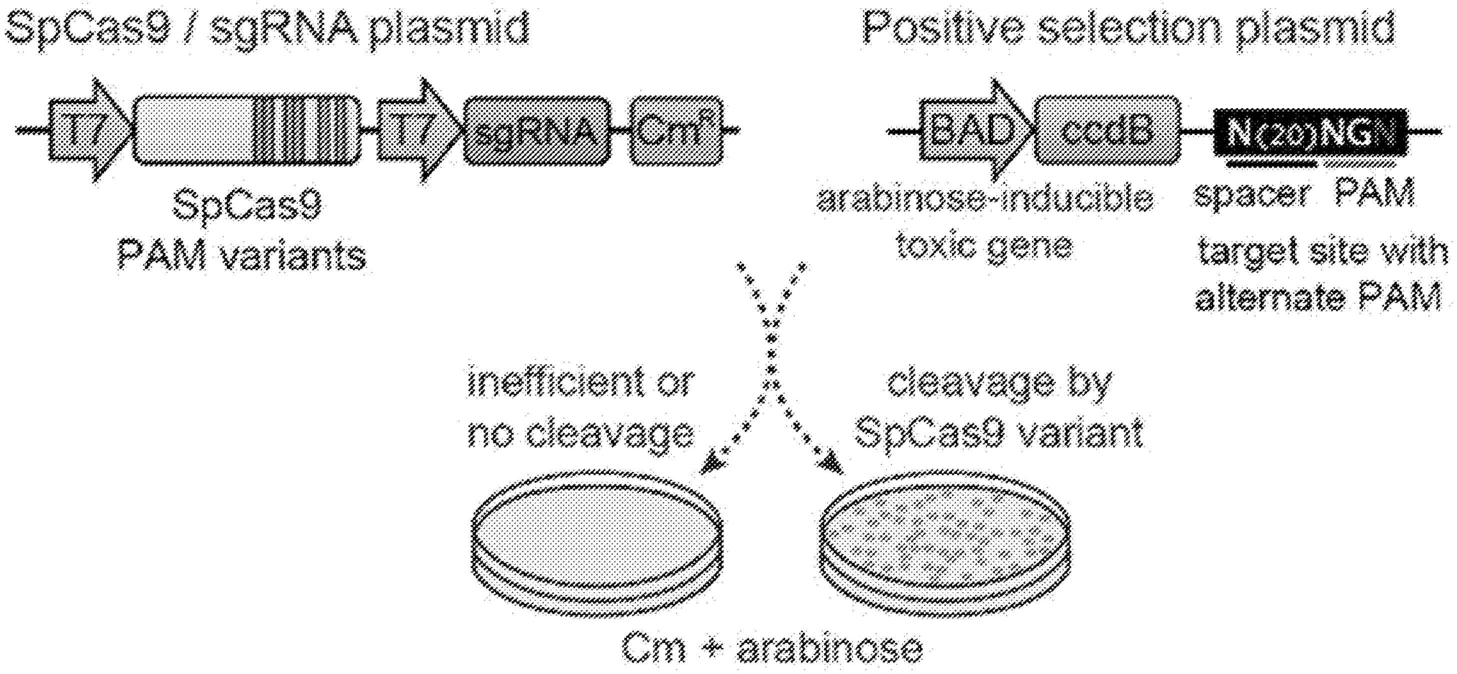
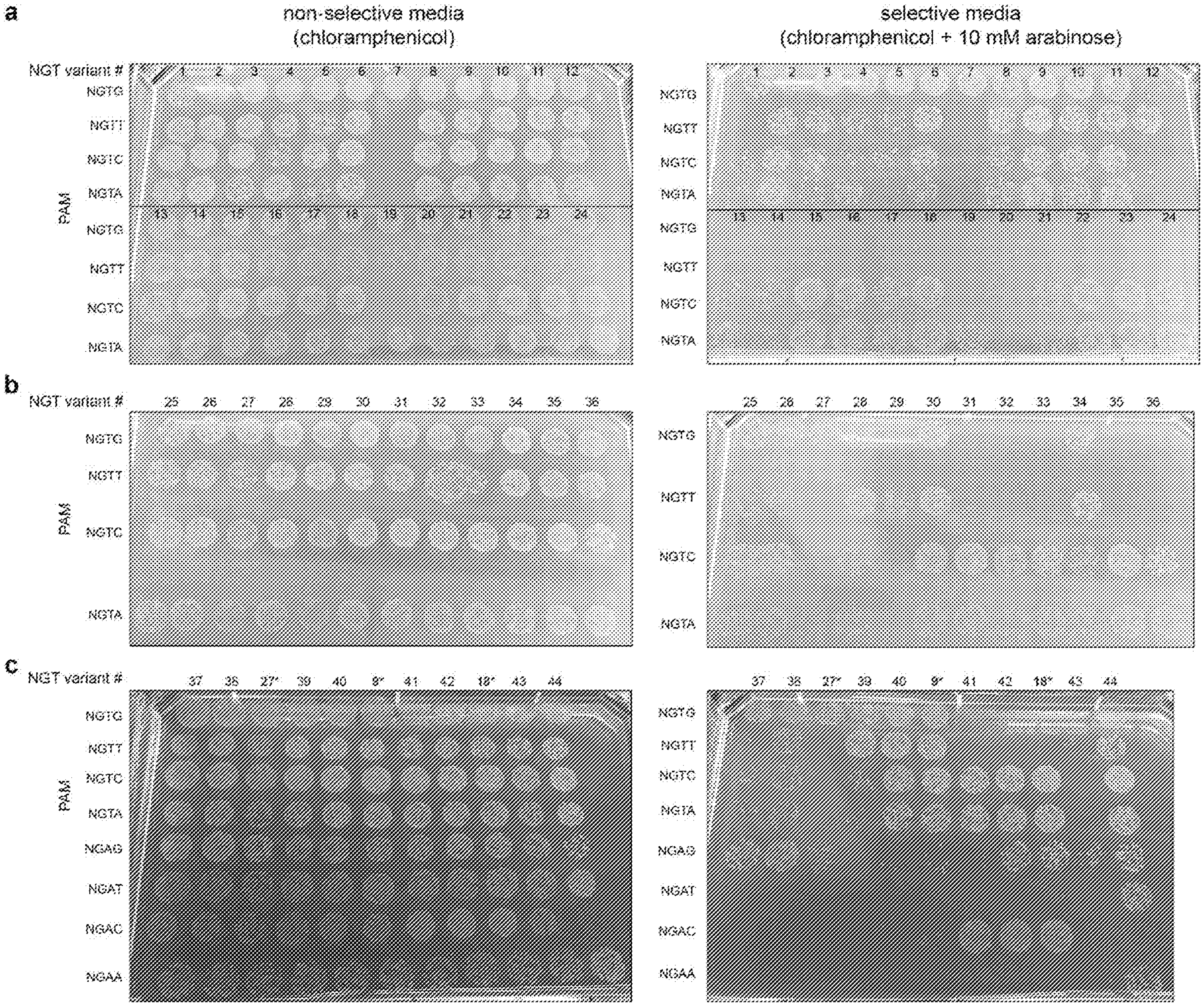
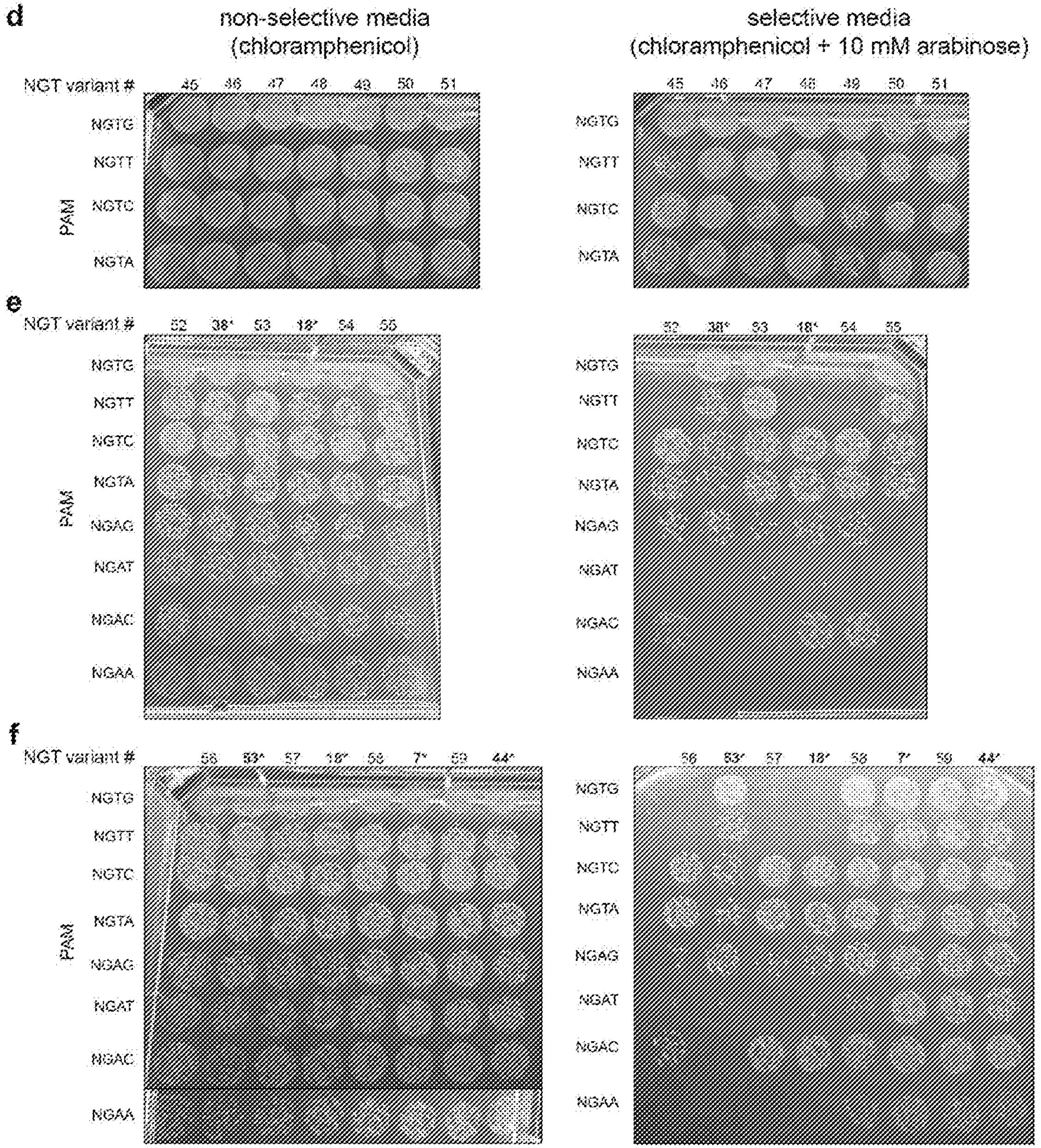

SpCas9 / sgRNA plasmid
T7
T7
sgRNA
Cm^R
SpCas9
PAM variants
Positive selection plasmid
BAD
ccdB
N(20)NGN
arabinose-inducible
toxic gene
spacer PAM
target site with
alternate PAM
inefficient or
no cleavage
cleavage by
SpCas9 variant
Cm + arabinose

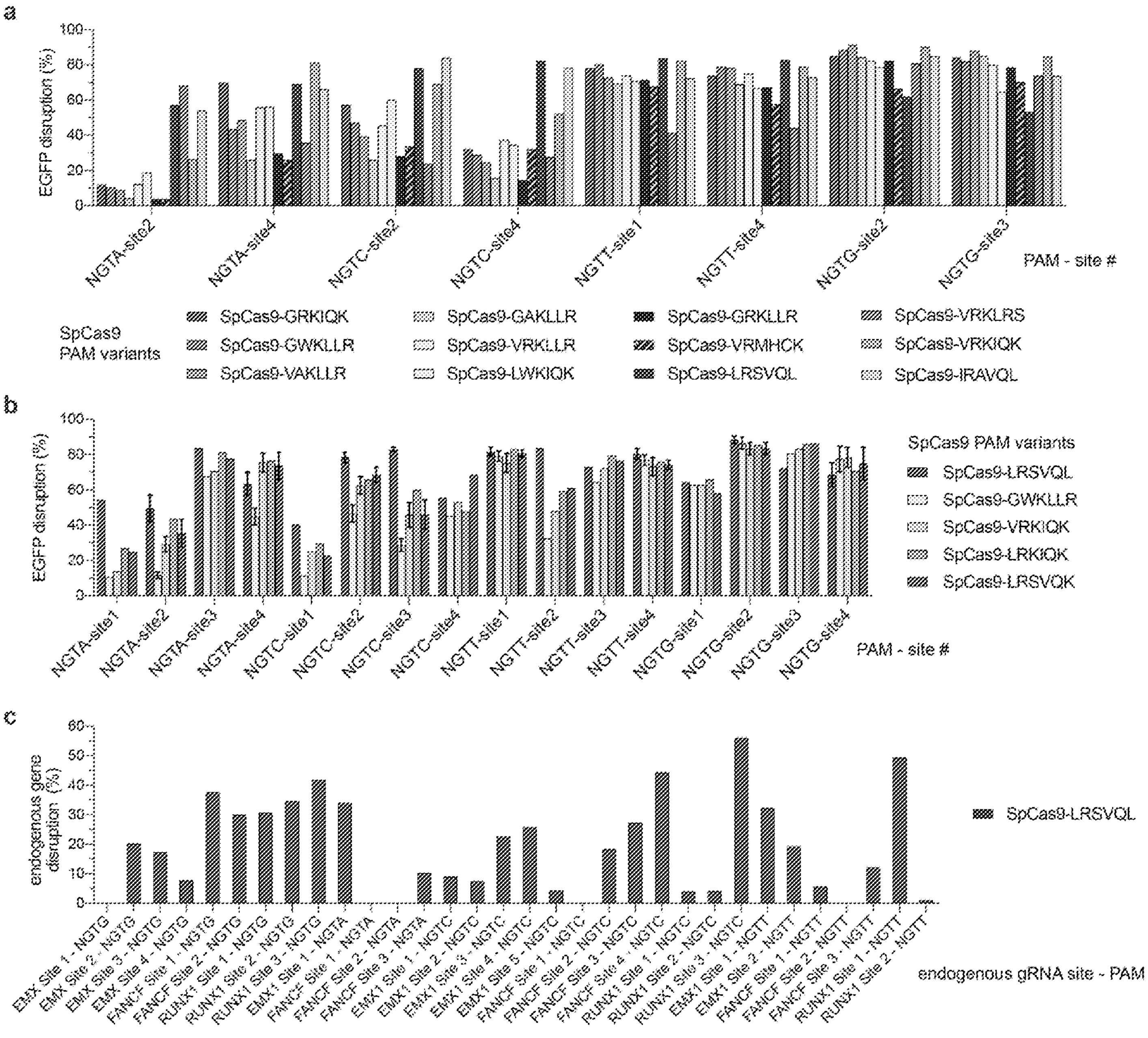
Fig. 1B


a
non-selective media
(chloramphenicol)
selective media
(chloramphenicol + 10 mM arabinose)
NGT variant #
1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18 19 20 21 22 23 24
PAM
NGTG
NGTT
NGTC
NGTA
b
25 26 27 28 29 30 31 32 33 34 35 36
c
37 38 27* 39 40 9* 41 42 18* 43 44
NGAG
NGAT
NGAC
NGAA

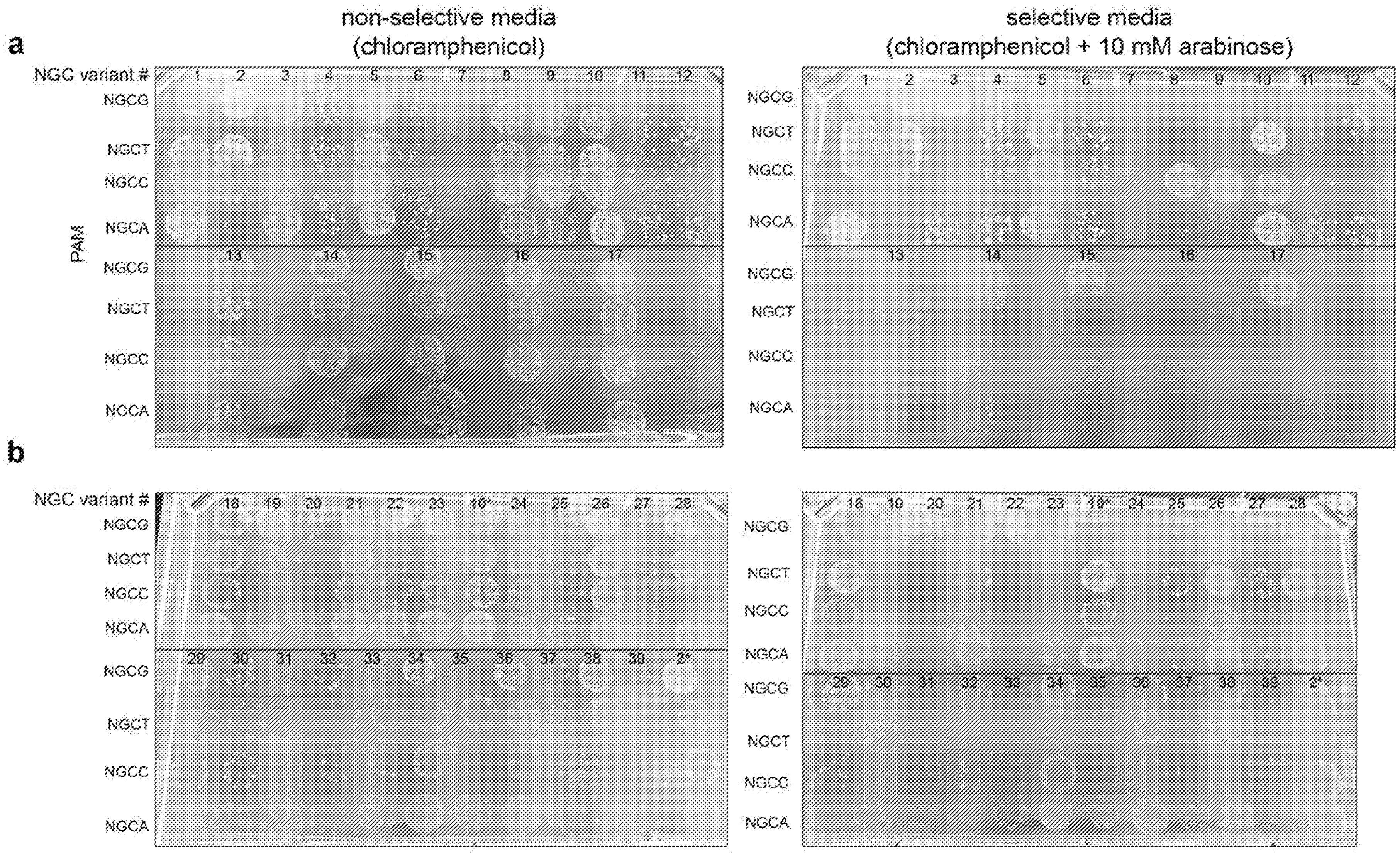
FIGs. 2A-C


d
non-selective media
(chloramphenicol)
selective media
(chloramphenicol + 10 mM arabinose)
NGT variant #
45 46 47 48 49 50 51
PAM
NGTG
NGTT
NGTC
NGTA
e
NGT variant #
52 38* 53 18* 54 55
NGAG
NGAT
NGAC
NGAA
f
NGT variant #
56 53* 57 18* 58 7* 59 44*

FIGs. 2D-F

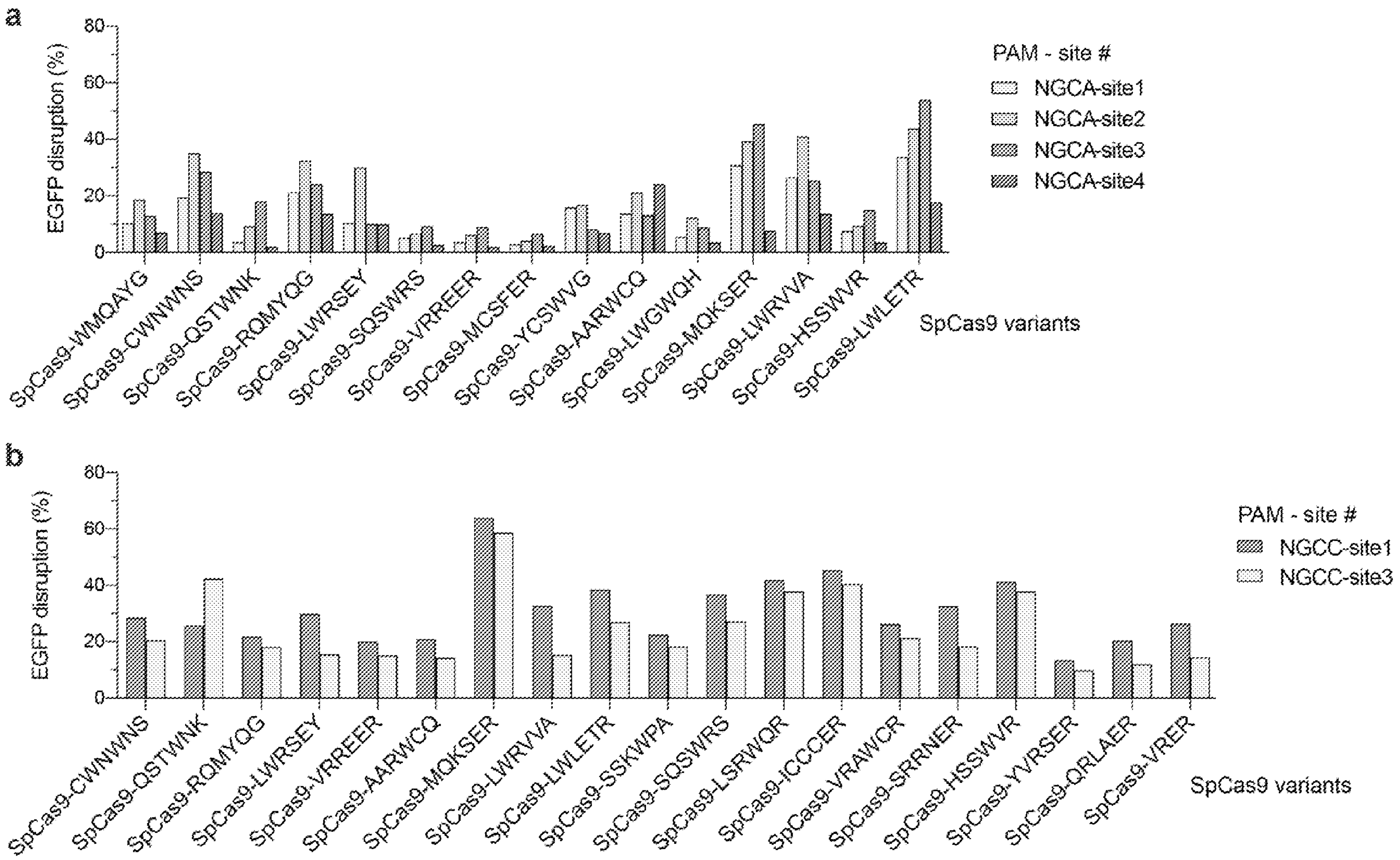

a
EGFP disruption (%)
0
20
40
60
80
100
NGTA-site2
NGTA-site4
NGTC-site2
NGTC-site4
NGTT-site1
NGTT-site4
NGTG-site2
NGTG-site3
PAM - site #
SpCas9 PAM variants
SpCas9-GRKIQK
SpCas9-GWKLLR
SpCas9-VAKLLR
SpCas9-GAKLLR
SpCas9-VRKLLR
SpCas9-LWKIQK
SpCas9-GRKLLR
SpCas9-VRMHCK
SpCas9-LRSVQL
SpCas9-VRKLRS
SpCas9-VRKIQK
SpCas9-IRAVQL
b
EGFP disruption (%)
0
20
40
60
80
100
NGTA-site1
NGTA-site2
NGTA-site3
NGTA-site4
NGTC-site1
NGTC-site2
NGTC-site3
NGTC-site4
NGTT-site1
NGTT-site2
NGTT-site3
NGTT-site4
NGTG-site1
NGTG-site2
NGTG-site3
NGTG-site4
PAM - site #
SpCas9 PAM variants
SpCas9-LRSVQL
SpCas9-GWKLLR
SpCas9-VRKIQK
SpCas9-LRKIQK
SpCas9-LRSVQK
c
endogenous gene disruption (%)
0
10
20
30
40
50
60
EMX Site 1 - NGTG
EMX Site 2 - NGTG
EMX Site 3 - NGTG
EMX Site 4 - NGTG
FANCF Site 1 - NGTG
FANCF Site 2 - NGTG
RUNX1 Site 1 - NGTG
RUNX1 Site 2 - NGTG
RUNX1 Site 3 - NGTG
EMX1 Site 1 - NGTA
FANCF Site 1 - NGTA
FANCF Site 2 - NGTA
FANCF Site 3 - NGTA
EMX1 Site 1 - NGTC
EMX1 Site 2 - NGTC
EMX1 Site 3 - NGTC
EMX1 Site 4 - NGTC
EMX1 Site 5 - NGTC
FANCF Site 1 - NGTC
FANCF Site 2 - NGTC
FANCF Site 3 - NGTC
FANCF Site 4 - NGTC
RUNX1 Site 1 - NGTC
RUNX1 Site 2 - NGTC
RUNX1 Site 3 - NGTC
EMX1 Site 1 - NGTT
EMX1 Site 2 - NGTT
FANCF Site 1 - NGTT
FANCF Site 2 - NGTT
FANCF Site 3 - NGTT
RUNX1 Site 1 - NGTT
RUNX1 Site 2 - NGTT
endogenous gRNA site - PAM
SpCas9-LRSVQL

FIGs. 3A-C

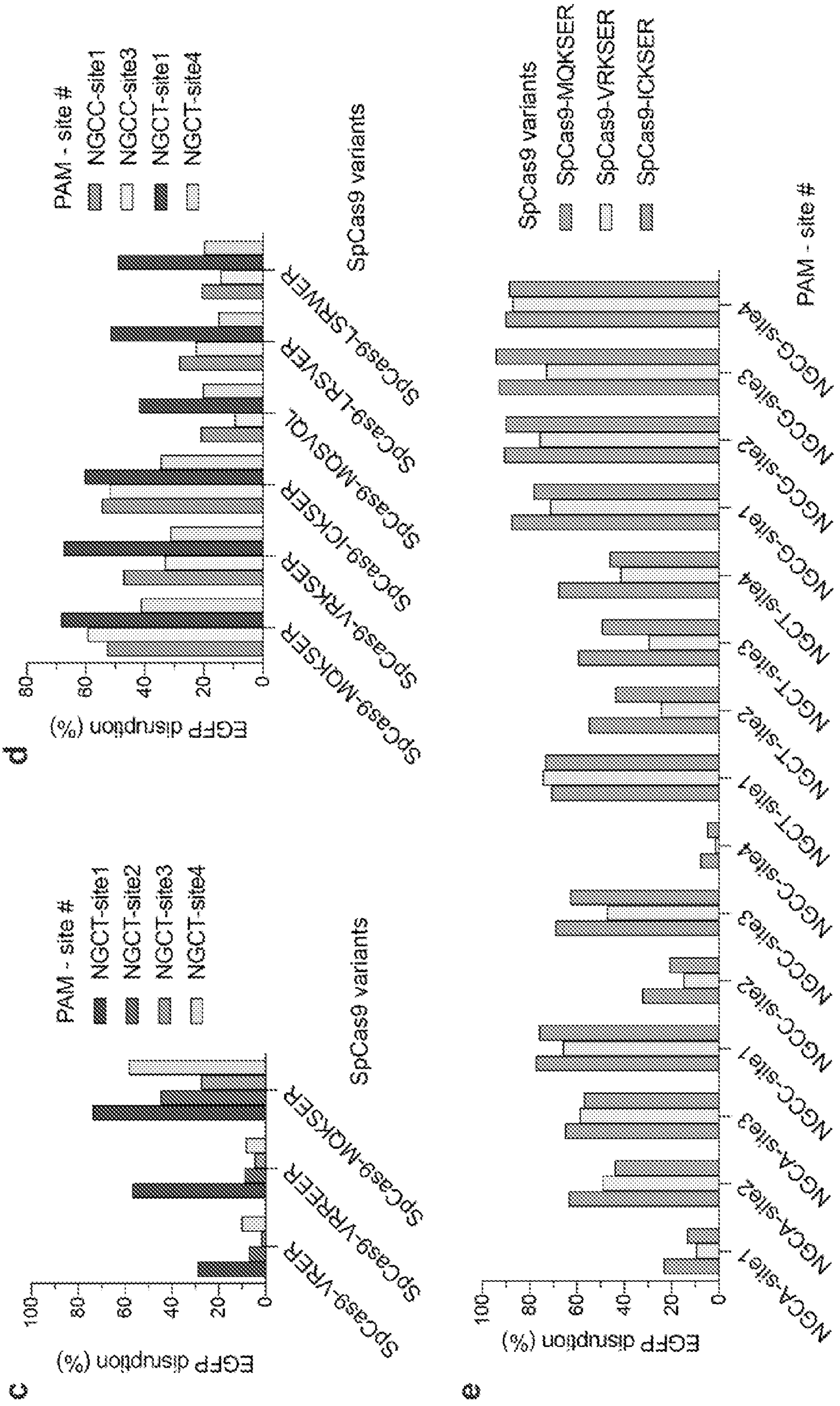

non-selective media
(chloramphenicol)

selective media
(chloramphenicol + 10 mM arabinose)

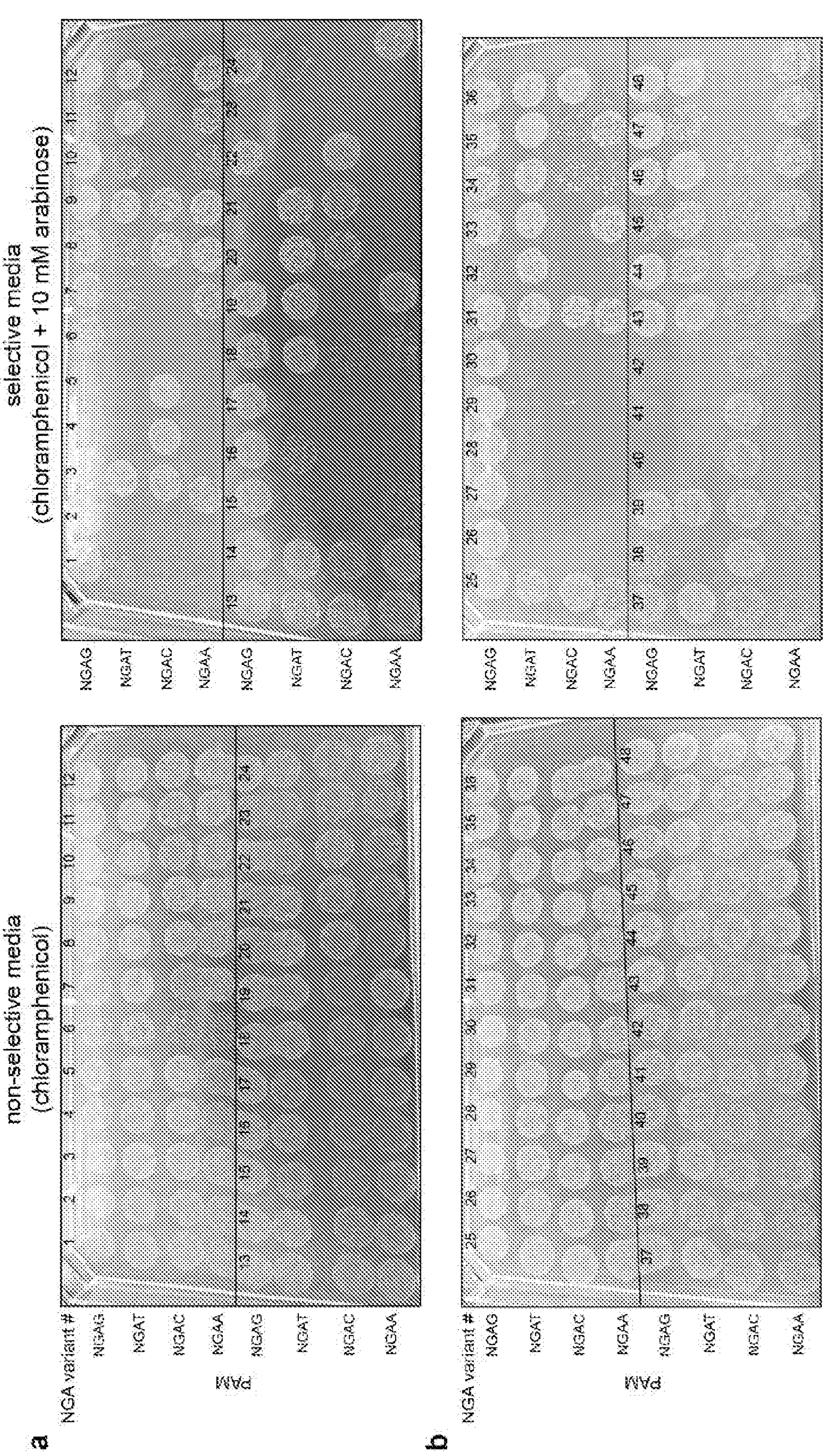
FIGs. 4A-B


a
EGFP disruption (%)
80
60
40
20
0
PAM - site #
NGCA-site1
NGCA-site2
NGCA-site3
NGCA-site4
SpCas9-WMQAYG
SpCas9-CWNWNS
SpCas9-QSTWNK
SpCas9-RQMYQG
SpCas9-LWRSEY
SpCas9-SQSWRS
SpCas9-VRREER
SpCas9-MCSFER
SpCas9-YCSWVG
SpCas9-AARWCQ
SpCas9-LWGWQH
SpCas9-MQKSER
SpCas9-LWRVVA
SpCas9-HSSWVR
SpCas9-LWLETR
SpCas9 variants
b
EGFP disruption (%)
80
60
40
20
0
PAM - site #
NGCC-site1
NGCC-site3
SpCas9-CWNWNS
SpCas9-QSTWNK
SpCas9-RQMYQG
SpCas9-LWRSEY
SpCas9-VRREER
SpCas9-AARWCQ
SpCas9-MQKSER
SpCas9-LWRVVA
SpCas9-LWLETR
SpCas9-SSKWPA
SpCas9-SQSWRS
SpCas9-LSRWQR
SpCas9-ICCCER
SpCas9-VRAWCR
SpCas9-SRRNER
SpCas9-HSSWVR
SpCas9-YVRSER
SpCas9-QRLAER
SpCas9-VRER
SpCas9 variants

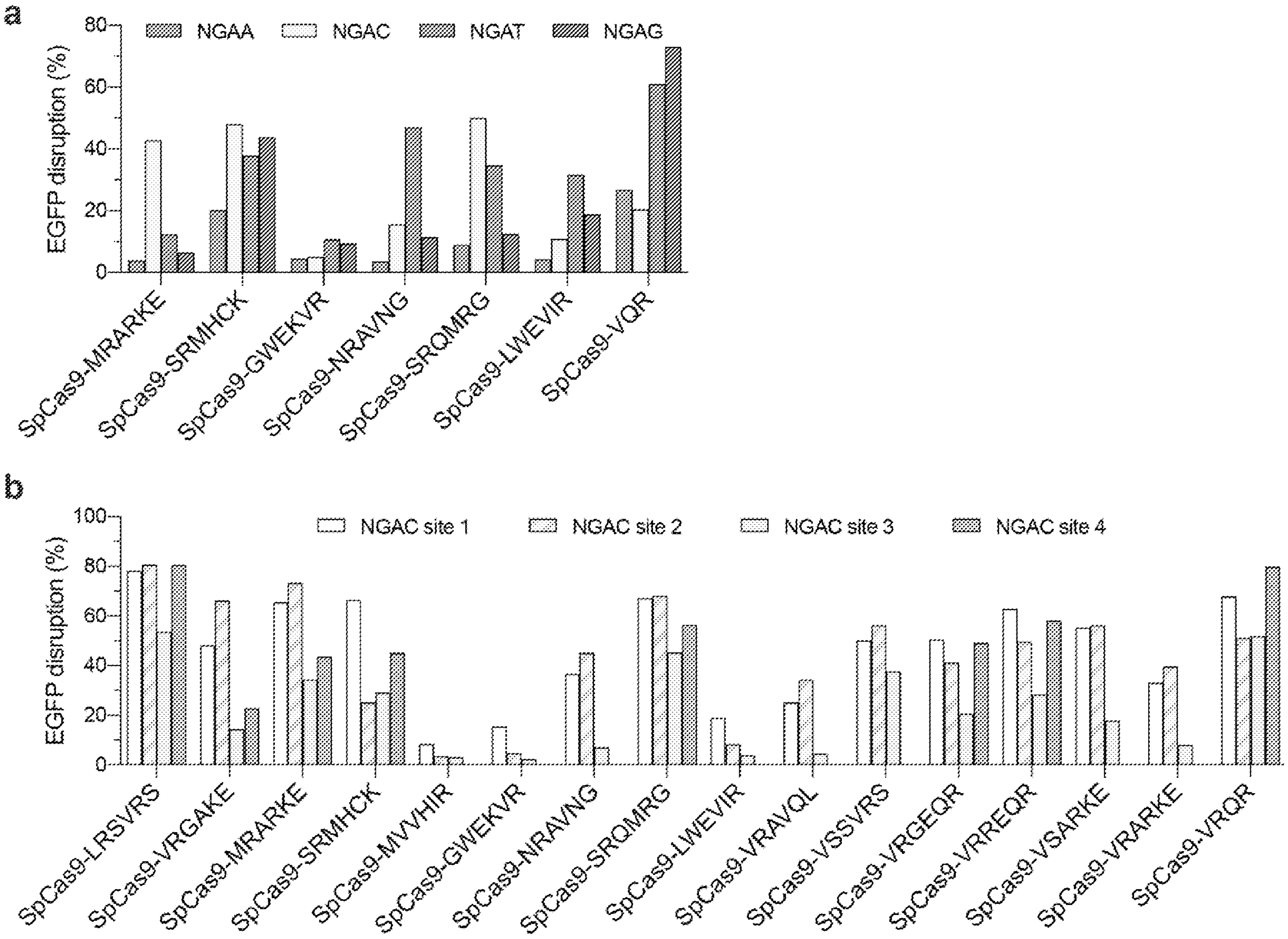
FIGs. 5A-B


c
EGFP disruption (%)
100
80
60
40
20
0
SpCas9-VRER
SpCas9-VRREER
SpCas9-MQKSER
SpCas9 variants
PAM - site #
NGCT-site1
NGCT-site2
NGCT-site3
NGCT-site4
d
EGFP disruption (%)
80
60
40
20
0
SpCas9-MQKSER
SpCas9-VRKSER
SpCas9-ICKSER
SpCas9-MQSVQL
SpCas9-LRSVER
SpCas9-LSRWER
SpCas9 variants
PAM - site #
NGCC-site1
NGCC-site3
NGCT-site1
NGCT-site4
e
EGFP disruption (%)
100
80
60
40
20
0
NGCA-site1
NGCA-site2
NGCA-site3
NGCC-site1
NGCC-site2
NGCC-site3
NGCC-site4
NGCT-site1
NGCT-site2
NGCT-site3
NGCT-site4
NGCG-site1
NGCG-site2
NGCG-site3
NGCG-site4
PAM - site #
SpCas9 variants
SpCas9-MQKSER
SpCas9-VRKSER
SpCas9-ICKSER

FIGs. 5C-E


a
b
non-selective media
(chloramphenicol)
selective media
(chloramphenicol + 10 mM arabinose)
NGA variant #
PAM
NGAG
NGAT
NGAC
NGAA
NGAG
NGAT
NGAC
NGAA
1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18 19 20 21 22 23 24
25 26 27 28 29 30 31 32 33 34 35 36
37 38 39 40 41 42 43 44 45 46 47 48

FIGs. 6A-B


a
EGFP disruption (%)
0
20
40
60
80
NGAA
NGAC
NGAT
NGAG
SpCas9-MRARKE
SpCas9-SRMHCK
SpCas9-GWEKVR
SpCas9-NRAVNG
SpCas9-SRQMRG
SpCas9-LWEVIR
SpCas9-VQR
b
EGFP disruption (%)
0
20
40
60
80
100
NGAC site 1
NGAC site 2
NGAC site 3
NGAC site 4
SpCas9-LRSVRS
SpCas9-VRGAKE
SpCas9-MRARKE
SpCas9-SRMHCK
SpCas9-MVVHIR
SpCas9-GWEKVR
SpCas9-NRAVNG
SpCas9-SRQMRG
SpCas9-LWEVIR
SpCas9-VRAVQL
SpCas9-VSSVRS
SpCas9-VRGEQR
SpCas9-VRREQR
SpCas9-VSARKE
SpCas9-VRARKE
SpCas9-VRQR

FIGs. 7A-B

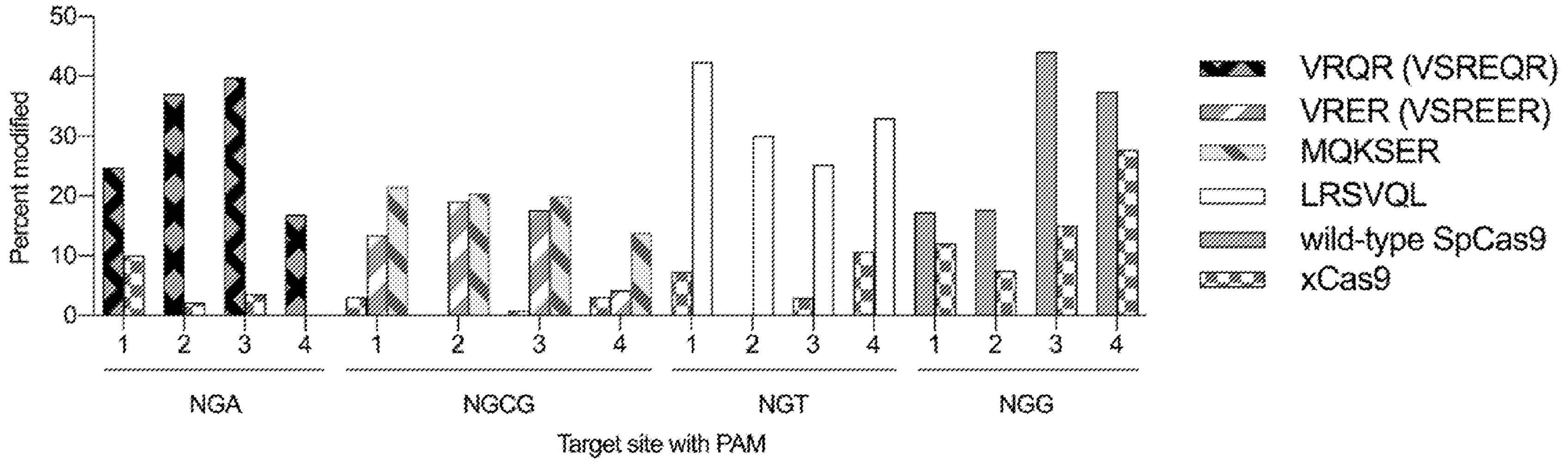

50
40
30
20
10
0
Percent modified
1
2
3
4
NGA
NGCG
NGT
NGG
Target site with PAM
VRQR (VSREQR)
VRER (VSREER)
MQKSER
LRSVQL
wild-type SpCas9
xCas9

FIG. 8

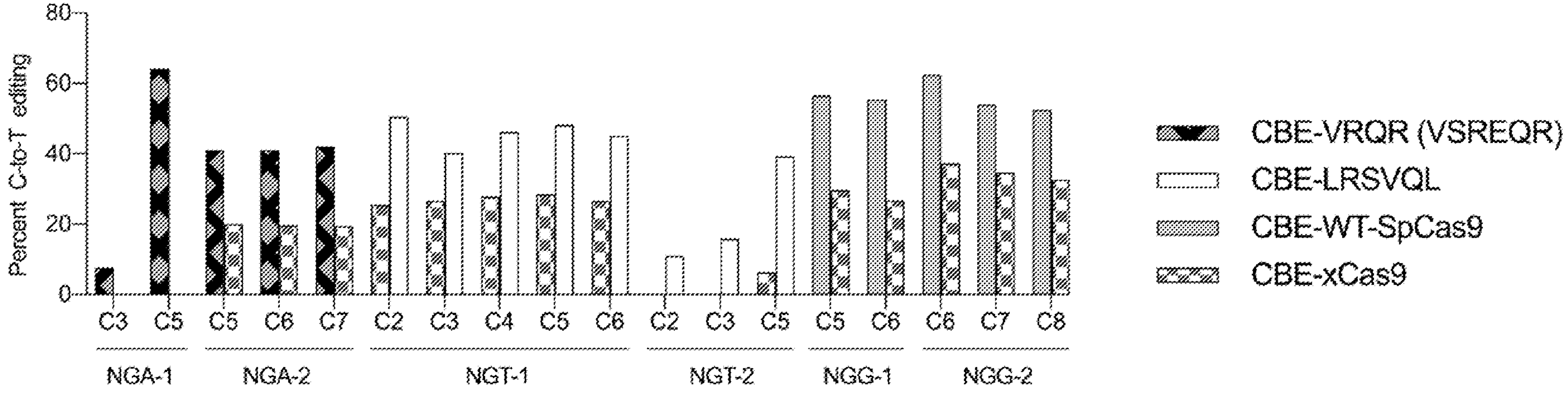

80
60
40
20
0
Percent C-to-T editing
C3
C5
C6
C7
C2
C4
C8
NGA-1
NGA-2
NGT-1
NGT-2
NGG-1
NGG-2
CBE-VRQR (VSREQR)
CBE-LRSVQL
CBE-WT-SpCas9
CBE-xCas9

FIG. 9A

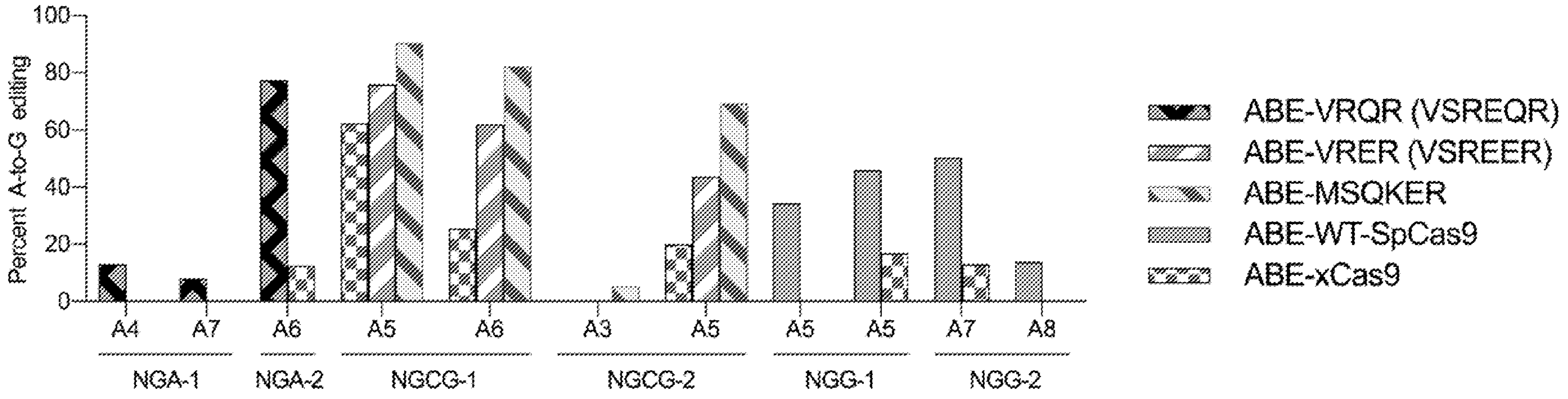

Percent A-to-G editing
100
80
60
40
20
0
A4
A7
A6
A5
A6
A3
A5
A5
A5
A7
A8
NGA-1
NGA-2
NGCG-1
NGCG-2
NGG-1
NGG-2
ABE-VRQR (VSREQR)
ABE-VRER (VSREER)
ABE-MSQKER
ABE-WT-SpCas9
ABE-xCas9

FIG. 9B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/047577

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 38/00; C07K 14/195; C07K 14/315; C07K 19/00; C12N; C12N 5/10; C12N 9/14 (2019.01)

CPC - A01K 2227/40; A61K 38/00; C07K 2319/71; C07K 2319/80; C12N 9/22; C12N 15/102; C12N 15/63; C12N 15/90; C12N 2800/22; C12N 2800/80; C12Y 301/00 (2019.01)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 424/94.6; 435/462 (keyword delimited)

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105543195 A (CHINA RICE RESEARCH INSTITUTE) 04 May 2016 (04.05.2016) entire document; see machine translation	1-4
A	KLEINSTIVER et al. "Engineered CRISPR-Cas9 nucleases with altered PAM specificities," Nature, 23 July 2015 (23.07.2015), Vol. 523, Pgs. 481-485. entire document	1-4
A	WO 2017/070633 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE et al) 27 April 2017 (27.04.2017) entire document	1-4
A	SLAYMAKER et al. "Rationally engineered Cas9 nucleases with improved specificity," Science, 01 December 2015 (01.12.2015), Vol. 351, Iss. 6268, Pgs. 84-88. entire document	1-4
A	WO 2016/141224 A1 (THE GENERAL HOSPITAL CORPORATION) 09 September 2016 (09.09.2016) entire document	1-4
T	NISHIMASU et al. "Engineered CRISPR-Cas9 nuclease with expanded targeting space," Science, 30 August 2018 (30.08.2018), Vol. 361, Iss. 6408, Pgs. 1259-1262. entire document	1-4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 08 January 2019	Date of mailing of the international search report 29 JAN 2019
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/047577

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

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

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

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

3. ☒ Claims Nos.: 0 36
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:

See extra sheet(s).

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-4 to the extent that they read on a mutant SpCas9 protein of SEQ ID NO:1 comprising the amino acid substitutions D1135L, S1136R, G1218S, E1219V, and T1337S.

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.

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-8 are drawn to Streptococcus pyogenes Cas9 (SpCas9) mutant proteins.

The first invention of Group I+ is restricted to a mutant Streptococcus pyogenes Cas9 (SpCas9) protein, wherein the mutant SpCas9 protein is selected to be SEQ ID NO:1 comprising the amino acid substitutions D1135L, S1136R, G1218S, E1219V, and T1337S. It is believed that claims 1-4 read on this first named invention and thus these claims will be searched without fee to the extent that they read on a mutant SpCas9 protein of SEQ ID NO:1 comprising the amino acid substitutions D1135L, S1136R, G1218S, E1219V, and T1337S.

Applicant is invited to elect additional mutant SpCas9 proteins, each with specified amino acid substitutions relative to SEQ ID NO:1, to be searched in a specific combination by paying an additional fee for each set of election. An exemplary election would be a mutant Streptococcus pyogenes Cas9 (SpCas9) protein, wherein the mutant SpCas9 protein is selected to be SEQ ID NO:1 comprising the amino acid substitutions D1135G, S1136R, G1218K, E1219I, R1335Q, and T1337K. Additional mutant SpCas9 proteins will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The Groups I+ formulas do not share a significant structural element responsible for altering a double stranded DNA (dsDNA) molecule in the genome of a cell, requiring the selection of alternatives for the mutant Streptococcus pyogenes Cas9 (SpCas9) protein, where "[t] he isolated protein ...comprising a set of mutations shown in Table A, 1, 2, or 3" and "The isolated protein of claims 1-3, comprising the following sets of mutations: LRSVQL, LRKIQK, LRSVQK, LWKIQK, VRKIQK, LWKIQK, IRAVQL, VRKLRS, GRKIQK, SWRVVV, SWKVLK, TAHFKV, MSGVKC, LRSVRS, SKTLRP, MWVHLN, TWSMRG, KRRCKV, VRAVQL, VSSVRS, VRSVRS, SRMHCK, GWKLLR, GWKOQK, VAKLLR, VAKIQK, VAKILR, GRKILR, VRKLLR, IRAVQL, VRKIQK, or VRMHCK variant; MQKSER, VRKSER, ICKSER, LRSVER, LWLETR, LSRWER, MQSVQL, VRREER, ICCCER, LSRWQR, LWRVVA, WMQAYG, LWRSEY, SQSWRS, LKAWRS, LWGWQH, MCSFER, LWMREQ, LWRVVA, HSSWVR, MWSEPT, GSNYQS, FMQWVN, YCSWVG, MCAWCG, FMQWVR, or SSKWPA; LRSVRS, SRQMRG, MRARKE, SRMHCK, VRREQR, VRGEQR, LRLSAR, AWTEVTR, KWMMCG, VRGAKE, AWNFQV, LWTTLN, SRMHCK, CWCQCV, AEEQQR, GWEKVR, NRAVNG, LRSYLH, VRGNNR, VQDAQR, GWRQSK, AWLCLS, KWARVV, MWAARP, SRMHCK, VKMAKG, QRKTRE, LCRQQR, CWSHQR, SRTHTQ, LWEVIR, VSSVRS, VRSVRS, IRAVRS, SRSVRS, LWKIQK, or VRMHCK variant".

Additionally, even if Groups I+ were considered to share the technical features of an isolated Streptococcus pyogenes Cas9 (SpCas9) protein, with mutations at three, four, five, or all six of the following positions: D1135, S1136, G1218, E1219, R1335, and/or T1337, wherein the mutations are not VSREER or VSREQR; these shared technical features do not represent a contribution over the prior art.

Specifically, CN 105543195 A to China Rice Research Institute discloses an isolated Cas9 protein ([t]he invention discloses plant Cas9 variant protein VQR, Abstract) with mutations at three of the following positions: D1135V, R1335Q, and T1337R (SEQ ID NO: 2 VQR corresponding protein sequence, Para. [0050]; where SEQ ID NO:2 of CN 105543195 A is the same as SEQ ID NO:1 of the instant application except for comprising the mutations D1135V, R1335Q, and T1337R [SEQ ID NO:2 of CN 105543195 A is provided on Pgs. 14-17 of the untranslated document).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.