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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 February 2018 (15.02.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/031683 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *A61K 48/00* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2017/046144

(22) International Filing Date:
09 August 2017 (09.08.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/372,755 09 August 2016 (09.08.2016) US
62/456,048 07 February 2017 (07.02.2017) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **LIU, David, R.**; 3 Whitman Circle, Lexington, MA 02420 (US). **CHAIKIND, Brian**; 91b Monte Cresta Avenue, Oakland, CA 94611 (US). **BESSEN, Jeffrey, L.**; 255 Pearl Street, Apt. #1, Cambridge, MA 02139 (US).

(74) Agent: **BAKER, C., Hunter**; Wolf, Greenfield & Sacks, P.C., 600 Atlanctic Avenue, Boston, MA 02210-2206 (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).

(54) Title: PROGRAMMABLE CAS9-RECOMBINASE FUSION PROTEINS AND USES THEREOF

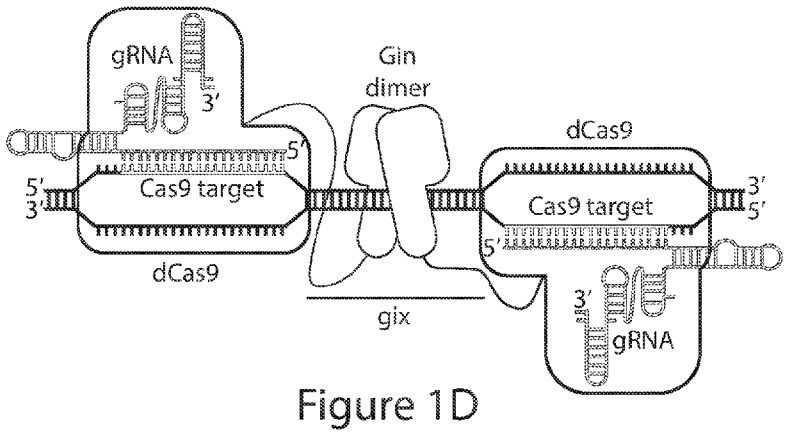


Figure 1D

(57) Abstract: Some aspects of this disclosure provide a fusion protein comprising a guide nucleotide sequence-programmable DNA binding protein domain (e.g., a nuclease-inactive variant of Cas9 such as dCas9), an optional linker, and a recombinase catalytic domain (e.g., a tyrosine recombinase catalytic domain or a serine recombinase catalytic domain such as a Gin recombinase catalytic domain). This fusion protein can recombine DNA sites containing a minimal recombinase core site flanked by guide RNA-specified sequences. The instant disclosure represents a step toward programmable, scarless genome editing in unmodified cells that is independent of endogenous cellular machinery or cell state.

WO 2018/031683 A1

[Continued on next page]

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

PROGRAMMABLE CAS9-RECOMBINASE FUSION PROTEINS AND USES THEREOF

RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. provisional patent application, U.S.S.N. 62/372,755, filed August 9, 2016, and U.S. provisional patent application, U.S.S.N. 62/456,048, filed February 7, 2017, each of which is incorporated herein by reference.

GOVERNMENT FUNDING

[0002] This invention was made with government support under grant numbers R01EB022376 and R35GM118062 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] Efficient, programmable, and site-specific homologous recombination remains a longstanding goal of genetics and genome editing. Early attempts at directing recombination to loci of interest relied on the transfection of donor DNA with long flanking sequences that are homologous to a target locus. This strategy was hampered by very low efficiency and thus the need for a stringent selection to identify integrants. More recent efforts have exploited the ability of double-stranded DNA breaks (DSBs) to induce homology-directed repair (HDR). Homing endonucleases and later programmable endonucleases such as zinc finger nucleases, TALE nucleases, Cas9, and fCas9 have been used to introduce targeted DSBs and induce HDR in the presence of donor DNA. In most post-mitotic cells, however, DSB-induced HDR is strongly down regulated and generally inefficient. Moreover, repair of DSBs by error-prone repair pathways such as non-homologous end-joining (NHEJ) or single-strand annealing (SSA) causes random insertions or deletions (indels) of nucleotides at the DSB site at a higher frequency than HDR. The efficiency of HDR can be increased if cells are subjected to conditions forcing cell-cycle synchronization or if the enzymes involved in NHEJ are inhibited. However, such conditions can cause many random and unpredictable events, limiting potential applications. The instant disclosure provides a fusion protein that can recombine DNA sites containing a minimal recombinase core site flanked by guide RNA-specified sequences and represents a step

toward programmable, scarless genome editing in unmodified cells that is independent of endogenous cellular machinery or cell state.

SUMMARY OF THE INVENTION

[0004] The instant disclosure describes the development of a fusion protein comprising a guide nucleotide sequence-programmable DNA binding protein domain, an optional linker, and a recombinase catalytic domain (*e.g.*, a serine recombinase catalytic domain such as a Gin recombinase catalytic domain, a tyrosine recombinase catalytic domain, or any evolved recombinase catalytic domain). This fusion protein operates on a minimal gix core recombinase site (NNNNAAASSWWSSTTTNNNN, SEQ ID NO: 19) flanked by two guide RNA-specified DNA sequences. Recombination mediated by the described fusion protein is dependent on both guide RNAs, resulting in orthogonality among different guide nucleotide:fusion protein complexes, and functions efficiently in cultured human cells on DNA sequences matching those found in the human genome. The fusion protein of the disclosure can also operate directly on the genome of human cells (*e.g.*, cultured human cells), catalyzing a deletion, insertion, inversion, translocation, or recombination between two recCas9 psuedosites located approximately 14 kilobases apart. This work provides engineered enzymes that can catalyze gene insertion, deletion, inversion, or chromosomal translocation with user-defined, single base-pair resolution in unmodified genomes.

[0005] In one aspect, the instant disclosure provides a fusion protein comprising: (i) a guide nucleotide sequence-programmable DNA binding protein domain; (ii) an optional linker; and (iii) a recombinase catalytic domain such as any serine recombinase catalytic domain (including but not limited to a Gin, Sin, Tn3, Hin, β, γδ, or PhiC31 recombinase catalytic domain), any tyrosine recombinase domain (including, but not limited to a Cre or FLP recombinase catalytic domain), or any evolved recombinase catalytic domain.

[0006] The guide nucleotide sequence-programmable DNA binding protein domain may be selected from the group consisting of nuclease inactive Cas9 (dCas9) domains, nuclease inactive Cpf1 domains, nuclease inactive Argonaute domains, and variants thereof. In certain embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain is a nuclease inactive Cas9 (dCas9) domain. In certain embodiments, the amino acid sequence of the dCas9 domain comprises mutations corresponding to a D10A and/or H840A mutation in SEQ ID NO: 1. In another embodiment, the amino acid sequence of the dCas9 domain comprises a mutation corresponding to a D10A mutation in SEQ ID NO: 1 and a

mutation corresponding to an H840A mutation in SEQ ID NO: 1. In another embodiment, the amino acid sequence of the dCas9 domain further does not include the N-terminal methionine shown in SEQ ID NO: 1. In a certain embodiment, the amino acid sequence of the dCas9 domain comprises SEQ ID NO: 712. In one embodiment, the amino acid sequence of the dCas9 domain has a greater than 95% sequence identity with SEQ ID NO: 712. In one embodiment, the amino acid sequence of the dCas9 domain has a greater than 96, 97, 98, 99% or greater sequence identity with SEQ ID NO: 712. In some embodiments, the recombinase catalytic domain is a serine recombinase catalytic domain or a tyrosine recombinase catalytic domain.

[0007] In one embodiment, the amino acid sequence of the recombinase catalytic domain is a Gin recombinase catalytic domain. In some embodiments, the Gin recombinase catalytic domain comprises a mutation corresponding to one or more of the mutations selected from: a H106Y, I127L, I136R and/or G137F mutation in SEQ ID NO: 713. In an embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises mutations corresponding to two or more of the mutations selected from: a I127L, I136R and/or G137F mutation in SEQ ID NO: 713. In an embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises mutations corresponding to a I127L, I136R and G137F mutation in SEQ ID NO: 713. In another embodiment, the amino acid sequence of the Gin recombinase has been further mutated. In a specific embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises SEQ ID NO: 713.

[0008] In another embodiment, the amino acid sequence of the recombinase catalytic domain is a Hin recombinase, β recombinase, Sin recombinase, Tn3 recombinase, γδ recombinase, Cre recombinase; FLP recombinase; or a phiC31 recombinase catalytic domain.

[0009] In one embodiment, the amino acid sequence of the Cre recombinase is truncated. In another embodiment, the tyrosine recombinase catalytic domain is the 25 kDa carboxy-terminal domain of the Cre recombinase. In another embodiment, the Cre recombinase begins with amino acid R118, A127, E138, or R154 (preceded in each case by methionine). In one embodiment, the amino acid sequence of the recombinase has been further mutated. In certain embodiments, the recombinase catalytic domain is an evolved recombinase catalytic domain. In some embodiments, the amino acid sequence of the recombinase has been further mutated.

[0010] In some embodiments, the linker (*e.g.*, the first, second, or third linker) may have a length of about 0 angstroms to about 81 angstroms. The linker typically has a length of about 33 angstroms to about 81 angstroms. The linker may be peptidic, non-peptidic, or a

combination of both types of linkers. In certain embodiments, the linker is a peptide linker. In certain embodiments, the peptide linker comprises an XTEN linker SGSETPGTSESATPES (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9), an amino acid sequence comprising one or more repeats of the tri-peptide GGS, or any of the following amino acid sequences: VPFLLEPDNINGKTC (SEQ ID NO: 10), GSAGSAAGSGEF (SEQ ID NO: 11), SIVAQLSRPDPA (SEQ ID NO: 12), MKIIEQLPSA (SEQ ID NO: 13), VRHKLKRVGS (SEQ ID NO: 14), GHGTGSTGSGSS (SEQ ID NO: 15), MSRPDPA (SEQ ID NO: 16), or GGSM (SEQ ID NO: 17). In another embodiment, the peptide linker comprises one or more repeats of the tri-peptide GGS. In one embodiment, the peptide linker comprises from one to five repeats of the tri-peptide GGS. In another embodiment, the peptide linker comprises from six to ten repeats of the tri-peptide GGS. In a specific embodiment, the peptide linker comprises eight repeats of the tri-peptide GGS. In another embodiment, the peptide linker is from 18 to 27 amino acids long. In certain embodiments, the peptide linker is 24 amino acids long. In certain embodiments, the peptide linker has the amino acid sequence GGSGGSGGSGGSGGSGGSGGSGGS (SEQ ID NO: 183).

[0011] In certain embodiments, the linker is a non-peptide linker. In certain embodiments, the non-peptide linker comprises polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacryl amide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker. In certain embodiments, the alkyl linker has the formula: —NH—$(CH_2)_s$—C(O)—, wherein s is any integer between 1 and 100, inclusive. In certain embodiments, s is any integer from 1-20, inclusive.

[0012] In another embodiment, the fusion protein further comprises a nuclear localization signal (NLS) domain. In certain embodiments, the NLS domain is bound to the guide nucleotide sequence-programmable DNA binding protein domain or the recombinase catalytic domain via one or more second linkers.

[0013] In one embodiment, the fusion protein comprises the structure NH_2-[recombinase catalytic domain]-[optional linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional, second linker sequence]-[NLS domain]-COOH. In certain embodiments, the fusion protein has greater than 85%, 90%, 95%, 98%, or 99% sequence identity with the amino acid sequence shown in SEQ ID NO: 719. In a specific embodiment, the fusion protein comprises the amino acid sequence shown

in SEQ ID NO: 719. In one embodiment, the fusion protein consists of the amino acid sequence shown in SEQ ID NO: 719.

[0014] In another embodiment, the fusion protein further comprises one or more affinity tags. In one embodiment, the affinity tag is selected from the group consisting of a FLAG tag, a polyhistidine (poly-His) tag, a polyarginine (poly-Arg) tag, a Myc tag, and an HA tag. In an embodiment, the affinity tag is a FLAG tag. In a specific embodiment, the FLAG tag has the sequence PKKKRKV (SEQ ID NO: 702). In another embodiment, the one or more affinity tags are bound to the guide nucleotide sequence-programmable DNA binding protein domain, the recombinase catalytic domain, or the NLS domain via one or more third linkers. In certain embodiments, the third linker is a peptide linker.

[0015] The elements of the fusion protein described herein may be in any order, without limitation. In some embodiments, the fusion protein has the structure NH_2-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH, NH_2-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH, or NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH.

[0016] In some embodiments, the fusion protein has the structure NH_2-[optional affinity tag]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH, NH_2-[optional affinity tag]-[optional linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-COOH, or NH_2-[optional affinity tag]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH.

[0017] In a certain embodiment, the fusion protein has greater than 85%, 90%, 95%, 98%, or 99% sequence identity with the amino acid sequence shown in SEQ ID NO: 185. In a specific embodiment, the fusion protein has the amino acid sequence shown in SEQ ID NO: 185. In certain embodiments, the recombinase catalytic domain of the fusion protein has greater than 85%, 90%, 95%, 98%, or 99% sequence identity with the amino acid sequence shown in amino acids 1-142 of SEQ ID NO: 185, which is identical to the sequence shown in SEQ ID NO: 713. In certain embodiments, the dCas9 domain has greater than 90%, 95%, or 99% sequence identity with the amino acid sequence shown in amino acids 167-1533 of SEQ ID NO: 185, which is identical to the sequence shown in SEQ ID NO: 712. In certain embodiments, the fusion protein of the instant disclosure has greater than 90%, 95%, or 99% sequence identity with the amino acid sequence shown in amino acids 1-1544 of SEQ ID NO: 185, which is identical to the sequence shown in SEQ ID NO: 719. In one embodiment, the fusion protein is bound to a guide RNA (gRNA).

[0018] In one aspect, the instant disclosure provides a dimer of the fusion protein described herein. In certain embodiments, the dimer is bound to a target DNA molecule. In certain embodiments, each fusion protein of the dimer is bound to the same strand of the target DNA molecule. In certain embodiments, each fusion protein of the dimer is bound to an opposite strand of the target DNA molecule. In certain embodiments, the gRNAs of the dimer hybridize to gRNA binding sites flanking a recombinase site of the target DNA molecule. In certain embodiments, the recombinase site comprises a res, gix, hix, six, resH, LoxP, FTR, or att core, or related core sequence. In certain embodiments, the recombinase site comprises a gix core or gix-related core sequence. In further embodiments, the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 3 to 7 base pairs. In certain embodiments, the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 5 to 6 base pairs.

[0019] In certain embodiments, a first dimer binds to a second dimer thereby forming a tetramer of the fusion protein. In one aspect, the instant disclosure provides a tetramer of the fusion protein described herein. In certain embodiments, the tetramer is bound to a target DNA molecule. In certain embodiments, each dimer is bound to an opposite strand of DNA. In other embodiments, each dimer is bound to the same strand of DNA.

[0020] In another aspect, the instant disclosure provides methods for site-specific recombination between two DNA molecules, comprising: (a) contacting a first DNA with a first fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain binds a first gRNA that hybridizes to a first region of the first DNA; (b)

contacting the first DNA with a second fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the first DNA; (c) contacting a second DNA with a third fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the third fusion protein binds a third gRNA that hybridizes to a first region of the second DNA; and (d) contacting the second DNA with a fourth fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a second region of the second DNA; wherein the binding of the fusion proteins in steps (a)-(d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNAs are recombined, and wherein the first, second, third, and/or fourth fusion protein is any of the fusion proteins described herein.

[0021] In one embodiment, the first and second DNA molecules have different sequences. In another embodiment, the gRNAs of steps (a) and (b) hybridize to opposing strands of the first DNA, and the gRNAs of steps (c) and (d) hybridize to opposing strands of the second DNA. In another embodiment, wherein the gRNAs of steps (a) and (b); and/or the gRNAs of steps (c) and (d) hybridize to regions of their respective DNAs that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart. In certain embodiments, the gRNAs of steps (a) and (b), and/or the gRNAs of steps (c) and (d) hybridize to regions of their respective DNAs at gRNA binding sites that flank a recombinase site (see, for example, Figure 1D). In certain embodiments, the recombinase site comprises a res, gix, hix, six, resH, LoxP, FTR, or att core, or related core sequence. In certain embodiments, the recombinase site comprises a gix core or gix-related core sequence. In certain embodiments, the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 3 to 7 base pairs. In certain embodiments, the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 5 to 6 base pairs.

[0022] The method for site-specific recombination provided herein may also be used with a single DNA molecule. In one aspect, the instant disclosure provides a method for site-specific recombination between two regions of a single DNA molecule, comprising: (a) contacting the DNA with a first fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain binds a first gRNA that hybridizes to a first region of the DNA; (b) contacting the DNA with a second fusion protein, wherein the guide

nucleotide sequence-programmable DNA binding protein domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the DNA; (c) contacting the DNA with a third fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the third fusion protein binds a third gRNA that hybridizes to a third region of the DNA; and (d) contacting the DNA with a fourth fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a fourth region of the DNA; wherein the binding of the fusion proteins in steps (a)-(d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNA is recombined, and wherein the first, second, third, and/or fourth fusion protein is any of the fusion proteins described.

[0023] In certain embodiments, the two regions of the single DNA molecule that are recombined have different sequences. In another embodiment, the recombination results in the deletion of a region of the DNA molecule. In a specific embodiment, the region of the DNA molecule that is deleted is prone to cross-over events in meiosis. In one embodiment, the first and second gRNAs of steps (a)-(d) hybridize to the same strand of the DNA, and the third and fourth gRNAs of steps (a)-(d) hybridize to the opposing strand of the DNA. In another embodiment, the gRNAs of steps (a) and (b) hybridize to regions of the DNA that are no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart, and the gRNAs of steps (c) and (d) hybridize to regions of the DNA that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, no more than 90, or no more than 100 base pairs apart. In certain embodiments, the gRNAs of steps (a) and (b); and/or the gRNAs of steps (c) and (d) hybridize to gRNA binding sites flanking a recombinase site. In certain embodiments, the recombinase site comprises a res, gix, hix, six, resH, LoxP, FTR, or att core or related core sequence. In one embodiment, the recombinase site comprises a gix core or gix-related core sequence. In certain embodiments, the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 3 to 7 base pairs. In certain embodiments, the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 5 to 6 base pairs.

[0024] The DNA described herein may be in a cell. In certain embodiments, the cell is a eukaryotic cell. In certain embodiments, the cell is a plant cell. In certain embodiments, the cell is a prokaryotic cell. In some embodiments, the cell may be a mammalian cell. In

some embodiments, the cell may be a human cell. In certain embodiments, the cell is in a subject. In some embodiments, the subject may be a mammal. In certain embodiments, the subject is a human. In certain embodiments, the cell may be a plant cell.

[0025] In one aspect, the instant disclosure provides a polynucleotide encoding any of the fusion proteins disclosed herein. In certain embodiments, the instant disclosure provides a vector comprising the polynucleotide encoding any of the fusion proteins disclosed herein.

[0026] In another aspect, the instant disclosure provides a cell comprising a genetic construct for expressing any fusion protein disclosed herein.

[0027] In one aspect, the instant disclosure provides a kit comprising any fusion protein disclosed herein. In another aspect, the instant disclosure provides a kit comprising a polynucleotide encoding any fusion protein disclosed herein. In another aspect, the instant disclosure provides a kit comprising a vector for recombinant protein expression, wherein the vector comprises a polynucleotide encoding any fusion protein disclosed herein. In another aspect, the instant disclosure provides a kit comprising a cell that comprises a genetic construct for expressing any fusion protein disclosed herein. In one embodiment, the kit further comprises one or more gRNAs and/or vectors for expressing one or more gRNAs.

[0028] The details of certain embodiments of the invention are set forth in the Detailed Description of Certain Embodiments, as described below. Other features, objects, and advantages of the invention will be apparent from the Definitions, Examples, Figures, and Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] **Figures 1A-1D**. Overview of the experimental setup. Cells are transfected with (**Figure 1A**) guide RNA expression vector(s) under the control of an hU6 promoter, (**Figure 1B**) a recCas9 expression vector under the control of a CMV promoter, and (**Figure 1C**) a recCas9 reporter plasmid. Co-transfection of these components results in reassembly of guide RNA-programmed recCas9 at the target sites (**Figure 1D**). This will mediate deletion of the polyA terminator, allowing transcription of GFP. Guide RNA expression vectors and guide RNA sequences are abbreviated as gRNA.

[0030] **Figures 2A-2F**. Optimization of fusion linker lengths and target site spacer variants. A single target guide RNA expression vector, pHU6-NT1, or non-target vector pHU6-BC74 was used in these experiments. The sequences can be found in Tables 6-9. (**Figure 2A**) A portion of the target site is shown with guide RNA target sites in black with dashed underline and a gix core sequence site in black. The 5′ and 3′ sequences on either side

of the pseudo-gix sites are identical, but inverted, and are recognized by pHU6-NT1. The number of base pairs spacers separating the gix pseudo-site from the 5′ and 3′ binding sites is represented by an X and Y, respectively. This figure depicts SEQ ID NOs: 700 and 703, respectively. (**Figure 2B**) Z represents the number of GGS repeats connecting Ginβ to dCas9. recCas9 activity is assessed when X=Y for (**Figure 2C**) $(GGS)_2$ (SEQ ID NO: 182), (**Figure 2D**) $(GGS)_5$ (SEQ ID NO: 701), and (**Figure 2E**) $(GGS)_8$ (SEQ ID NO: 183) linkers connecting the Gin catalytic domain to the dCas9 domain. (**Figure 2F**) The activity of recCas9 on target sites composed of uneven base pair spacers (X≠Y) was determined; X=Y=6 is included for comparison. All experiments are performed in triplicate and background fluorescence is subtracted from these experiments. The percentage of eGFP-positive cells is of only those transfected (*i.e.*, expressing a constitutively expressed iRFP gene) and at least 6,000 live events are recorded for each experiment. Guide RNA expression vectors and guide RNA sequences are abbreviated as "gRNA". Values and error bars represent the mean and standard deviation, respectively, of three independent biological replicates.

[0031] **Figures 3A-3B**. The dependence of forward and reverse guide RNAs on recCas9 activity. (**Figure 3A**) A sequence found within *PCDH15* replaces the target site tested in Figures 1A-1D. Two offset sequences can be targeted by guide RNAs on both the 5′ and 3′ sides of a pseudo-gix core site. This figure depicts SEQ ID NOs: 704-705, respectively. (**Figure 3B**) recCas9 activity was measured by co-transfecting a recCas9 expression vector and reporter plasmid with all four guide RNA expression vector pairs and individual guide RNA vectors with off target (O.T.) guide RNA vectors. The off-target forward and reverse contained guide RNA sequences targeting CLTA and VEGF, respectively. Control experiments transfected with the reporter plasmid but without a target guide RNA are also shown. The results of reporter plasmid cotransfected with different guide RNA expression vectors, but without recCas9 expression vectors, are also shown. All experiments were performed in quadruplicate, and background fluorescence is not subtracted from these experiments. The percentage of eGFP-positive cells is of only those transfected (*i.e.*, expressing a constitutively expressed iRFP gene), and at least 6,000 live events are recorded for each experiment. Guide RNA expression vectors and guide RNA sequences are abbreviated as gRNA. Values and error bars represent the mean and standard deviation, respectively, of four independent biological replicates.

[0032] **Figures 4A-4D**. recCas9 can target multiple sequences identical to those in the human genome. (**Figure 4A**) The target sites shown in Figures 1A-1D are replaced by

sequences found within the human genome. See Table 6 for sequences. A recCas9 expression vector was cotransformed with all combinations of guide RNA vectors pairs and reporter plasmids. Off-target guide RNA vectors were also cotransformed with the recCas9 expression vector and reporter plasmids and contain guide RNA sequences targeting CLTA and VEGF (see, *e.g.*, Guilinger *et al.*, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature biotechnology, (2014), the entire contents of which is hereby incorporated by reference). The percentage of eGFP-positive cells reflects that of transfected (iRFP-positive) cells. At least 6,000 live events are recorded for each experiment. Values and error bars represent the mean and standard deviation, respectively, of at least three independent biological replicates. (**Figure 4B**) Transfection experiments were performed again, replacing the resistance marker in the recCas9 expression vector and pUC with SpecR. After cotransfection and incubation, episomal DNA was extracted, transformed into *E. coli* and selected for carbenicillin resistance. Colonies were then sequenced to determine (**Figure 4C**) the ratio of recombined to fully intact plasmids. (**Figure 4D**) Sequencing data from episomal extractions isolated from transfected cells. Columns and rows represent the transfection conditions. Each cell shows the percent of recombined plasmid and the ratio. The values shown reflect the mean and standard deviation of two independent biological replicates. The average difference between the mean and each replicate is shown as the error. Guide RNA expression vectors and guide RNA sequences are abbreviated as gRNA.

[0033] **Figures 5A-5D.** recCas9 mediates guide RNA- and recCas9-dependent deletion of genomic DNA in cultured human cells. (**Figure 5A**) Schematic showing predicted recCas9 target sites located within an intronic region of the *FAM19A2* locus of chromosome 12 and the positions of primers used for nested PCR. This figure depicts SEQ ID NOs: 706-709 from top to bottom and left to right, respectively. (**Figure 5B**) Representative results of nested genomic PCR of template from cells transfected with the indicated expression vectors (n=3 biological replicates; NTC = no template control). The asterisk indicates the position of the 1.3-kb predicted primary PCR product. Arrow indicates the predicted deletion product after the secondary PCR. Both panes are from the same gel but were cut to remove blank lanes. (**Figure 5C**) Sanger sequencing of PCR products resulting from nested genomic PCR of cells transfected with all four gRNA expression vectors, and the recCas9 expression vector matches the predicted post-recombination product. This figure depicts SEQ ID NOs: 710 and 711 from top to bottom, respectively. (**Figure 5D**)

Estimated minimum deletion efficiency of *FAM19A2* locus determined by limiting-dilution nested PCR. The values shown reflect the mean and standard deviation of three replicates.

[0034] **Figure 6.** Reporter plasmid construction. Golden Gate assembly was used to construct the reporter plasmids described in this work. All assemblies started with a common plasmid, pCALNL-EGFP-Esp3I, that was derived from pCALNL-EGFP and contained to Esp3I restriction sites. The fragments shown are flanked by Esp3I sites. Esp3I digestion creates a series of compatible, unique 4-base pair 5′ overhangs so that assembly occurs in the order shown. To assemble the target sites, Esp3I (ThermoFisher Scientific, Waltham, MA) and five fragments were added to a single reaction tube to allow for iterative cycles of Esp3I digestion and T7 ligation. Reactions were then digested with Plasmid-Safe-ATP-dependent DNAse (Epicentre, Madison, WI) to reduce background. Colonies were analyzed by colony PCR to identify PCR products that matched the expected full length 5 part assembly product; plasmid from these colonies was then sent for sanger sequencing. For the genomic reporters shown in Figure 4, fragments 1 and 2 as well as fragments 4 and 5 were combined into two gBlocks (IDT, Coralville, IA) fragments encoding the entire target site (not shown in the figure). Assembly was then completed as described above. Details for construction can be found in the methods for the supporting material. Oligonucleotides and gBLOCKS for creation of fragments can be found in Table 2.

[0035] **Figures 7A and 7B.** A Cre recombinase evolved to target a site in the Rosa locus of the human genome called "36C6" was fused to dCas9. This fusion was then used to recombine a plasmid-based reporter containing the Rosa target site in a guide-RNA dependent fashion. **Figure 7A** demonstrates the results of linker optimization using wild-type Cre and 36C6. A GinB construct, targeting its cognate reporter, is shown for reference. The 1x 2x, 5x, and 8x linkers shown are the number of GGS repeats in the linker. **Figure 7B** shows the results of a reversion analysis which demonstrated that making mutations to 36C6 fused to dCas9 could impact the relative guide dependence of the chimeric fusion. A GinB construct, targeting its cognate reporter, is shown for reference. GGS-36C6: 1x GGS linker; 2GGS-36C6 (using linker SEQ ID NO: 181): 2x GGS linker (using linker SEQ ID NO: 181).

[0036] **Figure 8.** PAMs were identified flanking the Rosa26 site in the human genome that could support dCas9 binding (see at top). Guide RNAs and a plasmid reporter were designed to test whether the endogenous protospacers could support dCas9-36C6 activity. A GinB construct, targeting the gix reporter, is shown for reference. Mix: equal parts mixture of all 5 linker variants between Cas9 and 36C6. The sequences correspond to SEQ ID NO: 769 (the nucleotide sequence) and 770 (the amino acid sequence).

[0037] **Figures 9A-9B.** Locations of various tested truncations of Cre recombinase are shown in **Figure 9A**. Truncated variants of Cre recombinase fused to dCas9 show both appreciable recombinase activity as well as a strict reliance on the presence of guide RNA in a Lox plasmid reporter system (**Figure 9B**). Wild type Cre fused to dCas9 is shown as a positive control.

DEFINITIONS

[0038] As used herein, the singular forms "a," "an," and "the" include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to "an agent" includes a single agent and a plurality of such agents.

[0039] Non-limiting, exemplary RNA-programmable DNA-binding proteins include Cas9 nucleases, Cas9 nickases, nuclease inactive Cas9 (dCas9), CasX, CasY, Cpf1, C2c1, C2c2, C2C3, and Argonaute. The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active or inactive DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). Cas9 has two cleavage domains, which cut specific DNA strands (*e.g.*, sense and antisense strands). Cas9 nickases can be generated that cut either strand (including, but not limited to D10A and H840A of spCas9). A Cas9 domain (*e.g.*, nuclease active Cas9, nuclease inactive Cas9, or Cas9 nickases) may be used without limitation in the fusion proteins and methods described herein. Further, any of the guide nucleotide sequence-programmable DNA binding proteins described herein may be useful as nickases.

[0040] A Cas9 nuclease is also referred to sometimes as a *casn1* nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements, and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc), and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves a linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3′-5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNA sequences. However, single guide RNAs

("sgRNA", or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, *e.g.,* Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.,* "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease has an inactive (*e.g.,* an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase. As one example, the Cas9 nuclease (*e.g.*, Cas9 nickase) may cleave the DNA strand that is bound to the gRNA. As another example, the Cas9 nuclease (*e.g.*, Cas9 nickase) may cleave the DNA strand that is not bound to the gRNA. In another embodiment, any of the guide nucleotide sequence-programmable DNA binding proteins may have an inactive (*e.g.,* an inactivated) DNA cleavage domain, that is, the guide nucleotide sequence-programmable DNA binding protein is a nickase. As one example, the guide nucleotide sequence-programmable DNA binding protein may cleave the DNA strand that is bound to the gRNA. As another example, the guide nucleotide sequence-programmable DNA binding protein may cleave the DNA strand that is not bound to the gRNA.

[0041] Additional exemplary Cas9 sequences may be found in International Publication No.: WO/2017/070633, published April 27, 2017, and entitled "Evolved Cas9 Proteins for Gene Editing."

[0042] A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease "dead" Cas9). In some embodiments, dCas9 corresponds to, or comprises in part or in whole, the amino acid set forth as SEQ ID NO: 1, below. In some embodiments, variants of dCas9 (*e.g.,* variants of SEQ ID NO: 1) are provided. For example, in some embodiments, variants having mutations other than D10A and H840A are provided, which *e.g.,* result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domains of Cas9 (*e.g.,* substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of dCas9 (*e.g.,* variants of SEQ ID NO: 1) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO: 1. In some embodiments, variants of dCas9 (*e.g.,* variants of SEQ ID NO: 1) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 1, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more.

dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 1)

[0043] Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.,* Jinek *et al., Science.* 337:816-821(2012); Qi *et al.,* "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (See *e.g.,* Jinek *et al., Science.* 337:816-821(2012); Qi *et al., Cell.* 28;152(5):1173-83 (2013)). In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9, or fragments thereof, are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.,* a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to the corresponding fragment of wild type Cas9. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, SEQ ID NO: 2 (nucleotide); SEQ ID NO: 3 (amino acid)). In some embodiments the Cas9 domain comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to wild type Cas9. In some embodiments, the Cas9 domain comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more or more mutations compared to wild type Cas9. In some embodiments, the Cas9 domain comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least

30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues as compared to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9. In some embodiments, the fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9.

[0044] In some embodiments, the fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or 1300 amino acids in length.

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATG
ATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCT
TATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGA
AGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAG
ATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCC
TATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAA
AAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGT
TTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCA
GTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCG
ATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGA
GAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGA
TTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCG
CAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAG
ATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATC
TTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATA
AATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGA
TTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCAT
GCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCT
TGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCG
GAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCA
TTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGC
TTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACC
AGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACC
GTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTG
AAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTT

GGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGG
ATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTC
GCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAA
AACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGAT
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA
TTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACT
GGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAA
AAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGA
TTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAA
TGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATT
GTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTG
GTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAA
CGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTT
GATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTT
TGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTT
AAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTAC
CATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTG
AATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGA
AATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACA
CTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGG
ATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAAC
AGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCT
CGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAG
TGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAAT
TATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAA
AAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTT
AGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAG
CATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATG
CCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAA
TATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATT
GATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTC
TTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO: 2)

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKA
ILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS
FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT
VKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRG
MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD
SLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 3)

[0045] In some embodiments, wild type Cas9 corresponds to, or comprises, SEQ ID NO: 4 (nucleotide) and/or SEQ ID NO: 5 (amino acid).

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCC
CTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC
AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAG
GTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAAT
CCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTG
ATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCA
AATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTC
GACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTT
ATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGA
AAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGG
CAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTAC
TATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
TGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAG
AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACG
AAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
CTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTC
GATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
ATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTC
TTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGT
GGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGAC
TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCG
AATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
ATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAA
AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCT
GTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGAT
TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTC
GTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTA
ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGC
CAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
CGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
AGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAA
TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATC
ACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
AAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
TGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAG
AAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
AAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAG
TATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAA
CTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAA
GATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAA
TTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAG
TATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAA
TCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGG
AAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC
AAGGCTGCAGGA (SEQ ID NO: 4)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO: 5)

[0046] In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1), *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); or *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) or to a Cas9 from any other organism.

[0047] Cas9 recognizes a short motif (PAM motif) in the CRISPR repeat sequences in the target DNA sequence. A "PAM motif," or "protospacer adjacent motif," as used herein, refers a DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. PAM is a component of the invading virus or plasmid, but is not a component of the bacterial CRISPR locus. Naturally, Cas9 will not successfully bind to or cleave the target DNA sequence if it is not followed by the PAM sequence. PAM is a targeting component (not found in the bacterial genome) which distinguishes bacterial self from non-self DNA, thereby preventing the CRISPR locus from being targeted and destroyed by the Cas9 nuclease activity.

[0048] Wild-type *Streptococcus pyogenes* Cas9 recognizes a canonical PAM sequence (*e.g.*, Cas9 from *Streptococcus thermophiles*, *Staphylococcus aureus*, *Neisseria meningitidis*, or *Treponema denticolaor*) and Cas9 variants thereof have been described in the art to have different, or more relaxed PAM requirements. Typically, Cas9 proteins, such as

Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the "N" in "NGG" is adenine (A), thymine (T), guanine (G), or cytosine (C), and the G is guanine. This may limit the ability to edit desired bases within a genome. In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is within a 4 base region (*e.g.,* a "deamination window"), which is approximately 15 bases upstream of the PAM. See Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016), the entire contents of which are hereby incorporated by reference. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base region. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM. Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical (*e.g.,* NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); the entire contents of each are hereby incorporated by reference. See also: Klenstiver *et al.*, *Nature* 529, 490–495, 2016; Ran *et al.*, *Nature*, Apr 9; 520(7546): 186–191, 2015; Hou *et al.*, *Proc Natl Acad Sci U S A*, 110(39):15644-9, 2014; Prykhozhij *et al.*, *PLoS One*, 10(3): e0119372, 2015; Zetsche *et al.*, *Cell* 163, 759–771, 2015; Gao *et al.*, *Nature Biotechnology*, doi:10.1038/nbt.3547, 2016; Want *et al.*, *Nature* 461, 754–761, 2009; Chavez *et al.*, doi: dx dot doi dot org/10.1101/058974; Fagerlund *et al.*, *Genome Biol.* 2015; 16: 25, 2015; Zetsche *et al.*, *Cell*, 163, 759–771, 2015; and Swarts *et al.*, *Nat Struct Mol Biol*, 21(9):743-53, 2014, the entire contents of each of which is incorporated herein by reference.

[0049] Thus, the guide nucleotide sequence-programmable DNA-binding protein of the present disclosure may recognize a variety of PAM sequences including, without limitation: NGG, NGAN (SEQ ID NO: 741), NGNG (SEQ ID NO: 742), NGAG (SEQ ID NO: 743), NGCG (SEQ ID NO: 744), NNGRRT (SEQ ID NO: 745), NGRRN (SEQ ID NO: 746), NNNRRT (SEQ ID NO: 747), NNNGATT (SEQ ID NO: 748), NNAGAAW (SEQ ID

NO: 749), NAAAC (SEQ ID NO: 750), TTN, TTTN (SEQ ID NO: 751), and YTN, wherein Y is a pyrimidine, and N is any nucleobase.

[0050] One example of an RNA-programmable DNA-binding protein that has different PAM specificity is Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1 (Cpf1). Similar to Cas9, Cpf1 is also a class 2 CRISPR effector. It has been shown that Cpf1mediates robust DNA interference with features distinct from Cas9. Cpf1 is a single RNA-guided endonuclease lacking tracrRNA, and it utilizes a T-rich protospacer-adjacent motif (TTN, TTTN (SEQ ID NO: 751), or YTN). Moreover, Cpf1 cleaves DNA via a staggered DNA double-stranded break. Out of 16 Cpf1-family proteins, two enzymes from *Acidaminococcus* and *Lachnospiraceae* are shown to have efficient genome-editing activity in human cells.

[0051] Also provided herein are nuclease-inactive Cpf1 (dCpf1) variants that may be used as a RNA-programmable DNA-binding protein domain. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9 but does not have a HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alpha-helical recognition lobe of Cas9. It was shown in Zetsche *et al.*, *Cell*, 163, 759–771, 2015 (the entire contents of which is incorporated herein by reference) that the RuvC-like domain of Cpf1 is responsible for cleaving both DNA strands and inactivation of the RuvC-like domain inactivates Cpf1 nuclease activity. For example, mutations corresponding to D917A, E1006A, or D1255A in *Francisella novicida* Cpf1 (SEQ ID NO: 714) inactivates Cpf1 nuclease activity. In some embodiments, the dCpf1 of the present disclosure comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/ E1006A/D1255A in SEQ ID NO: 714. It is to be understood that any mutations, *e.g.*, substitution mutations, deletions, or insertions that inactivates the RuvC domain of Cpf1 may be used in accordance with the present disclosure.

[0052] In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain of the present disclosure has no requirements for a PAM sequence. One example of such a guide nucleotide sequence-programmable DNA-binding protein may be an Argonaute protein from *Natronobacterium gregoryi* (NgAgo). NgAgo is a ssDNA-guided endonuclease. NgAgo binds 5′ phosphorylated ssDNA of ~24 nucleotides (gDNA) to guide it to its target site and will make DNA double-strand breaks at gDNA site. In contrast to Cas9, the NgAgo–gDNA system does not require a protospacer-adjacent motif (PAM). Using a nuclease inactive NgAgo (dNgAgo) can greatly expand the codons that may be targeted. The characterization and use of NgAgo have been described in Gao *et al.*, *Nat*

Biotechnol. Epub 2016 May 2. PubMed PMID: 27136078; Swarts *et al.*, *Nature*. 507(7491) (2014):258-61; and Swarts *et al.*, *Nucleic Acids Res.* 43(10) (2015):5120-9, the entire contents of each of which are incorporated herein by reference. The sequence of *Natronobacterium gregoryi* Argonaute is provided in SEQ ID NO: 718.

[0053] Also provided herein are Cas9 variants that have relaxed PAM requirements (PAMless Cas9). PAMless Cas9 exhibits an increased activity on a target sequence that does not comprise a canonical PAM (NGG) at its 3′-end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 1, *e.g.,* increased activity by at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold. Thus, the dCas9 or Cas9 nickase of the present disclosure may further comprise mutations that relax the PAM requirements, *e.g.,* mutations that correspond to A262T, K294R, S409I, E480K, E543D, M694I, or E1219V in SEQ ID NO: 1.

[0054] It should be appreciated that additional Cas9 proteins (*e.g.,* a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9), including variants and homologs thereof, are within the scope of this disclosure. Exemplary Cas9 proteins include, without limitation, those provided below. In some embodiments, the Cas9 protein is a nuclease dead Cas9 (dCas9). In some embodiments, the dCas9 comprises the amino acid sequence shown below. In some embodiments, the Cas9 protein is a Cas9 nickase (nCas9). In some embodiments, the nCas9 comprises the amino acid sequence shown below. In some embodiments, the Cas9 protein is a nuclease active Cas9. In some embodiments, the nuclease active Cas9 comprises the amino acid sequence shown below.

Exemplary catalytically inactive Cas9 (dCas9):
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS

VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:
752)

Exemplary Cas9 nickase (nCas9):
DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV

REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:
753)

Exemplary catalytically active Cas9:
DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA

PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 754)

[0055] In some embodiments, Cas9 refers to a Cas9 from arehaea (*e.g.* nanoarchaea), which constitute a domain and kingdom of single-celled prokaryotic microbes. In some embodiments, Cas9 refers to CasX or CasY, which have been described in, for example, Burstein et al., "New CRISPR–Cas systems from uncultivated microbes." *Cell Res.* 2017 Feb 21. doi: 10.1038/cr.2017.21, the entire contents of which is hereby incorporated by reference. Using genome-resolved metagenomics, a number of CRISPR–Cas systems were identified, including the first reported Cas9 in the archaeal domain of life. This divergent Cas9 protein was found in little-studied nanoarchaea as part of an active CRISPR–Cas system. In bacteria, two previously unknown systems were discovered, CRISPR–CasX and CRISPR–CasY, which are among the most compact systems yet discovered. In some embodiments, Cas9 refers to CasX, or a variant of CasX. In some embodiments, Cas9 refers to a CasY, or a variant of CasY. It should be appreciated that other RNA-guided DNA binding proteins may be used as a guide nucleotide sequence-programmable DNA-binding protein, and are within the scope of this disclosure.

[0056] In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain of any of the fusion proteins provided herein may be a CasX or CasY protein. In some embodiments, guide nucleotide sequence-programmable DNA-binding protein domain is a CasX protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain is a CasY protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to a naturally-occurring CasX or CasY protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain is a naturally-occurring CasX or CasY protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of of the exemplary CasX or CasY proteins described herein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain comprises an amino acid sequence of any one of of the exemplary CasX or CasY proteins described herein. It should

be appreciated that CasX and CasY from other bacterial species may also be used in accordance with the present disclosure.

CasX (uniprot.org/uniprot/F0NN87; uniprot.org/uniprot/F0NH53)
>tr|F0NN87|F0NN87_SULIH CRISPR-associated Casx protein OS=Sulfolobus islandicus (strain HVE10/4) GN=SiH_0402 PE=4 SV=1
MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQV
KECEEVSAPSFVKPEFYEFGRSPGMVERTRRVKLEVEPHYLIIAAAGWVLTRLGKAK
VSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVSV
VRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVL
ANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG
(SEQ ID NO: 755)

>tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS=Sulfolobus islandicus (strain REY15A) GN=SiRe_0771 PE=4 SV=1
MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAE
RRGKAKKKKGEEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQV
KECEEVSAPSFVKPEFYKFGRSPGMVERTRRVKLEVEPHYLIMAAAGWVLTRLGKA
KVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVVSSVTNPNVS
VVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIV
LANYIYEYLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG
(SEQ ID NO: 756)

CasY (ncbi.nlm.nih.gov/protein/APG80656.1)
>APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]
MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDD
YVGLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEV
RGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRE
RHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTC
CLLPFDTVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGE
GFLGRLRENKITELKKAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHW
GGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAV
VSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLE

AEKKKKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKN
AAIYTDALWKAVEKIYKSAFSSSLKNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWK
PIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVRLPSTENIAKAGIALARE
LSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTVKD
FMKTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELL
YIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGY
ELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLGRGQNKIVLYVRSSYYQTQ
FLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPF
TIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGL
KLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKI
KKVYATLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFCGACKKLWRAE
MQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHISKK
MRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQ
MKKI (SEQ ID NO: 757)

[0057] The terms "conjugating," "conjugated," and "conjugation" refer to an association of two entities, for example, of two molecules such as two proteins, two domains (*e.g.*, a binding domain and a cleavage domain), or a protein and an agent, *e.g.*, a protein binding domain and a small molecule. In some aspects, the association is between a protein (*e.g.*, RNA-programmable nuclease) and a nucleic acid (*e.g.*, a guide RNA). The association can be, for example, via a direct or indirect (*e.g.*, via a linker) covalent linkage. In some embodiments, the association is covalent. In some embodiments, two molecules are conjugated via a linker connecting both molecules. For example, in some embodiments where two proteins are conjugated to each other, *e.g.*, a binding domain and a cleavage domain of an engineered nuclease, to form a protein fusion, the two proteins may be conjugated via a polypeptide linker, *e.g.*, an amino acid sequence connecting the C-terminus of one protein to the N-terminus of the other protein.

[0058] The term "consensus sequence," as used herein in the context of nucleic acid sequences, refers to a calculated sequence representing the most frequent nucleotide residues found at each position in a plurality of similar sequences. Typically, a consensus sequence is determined by sequence alignment in which similar sequences are compared to each other and similar sequence motifs are calculated. In the context of recombinase target site sequences, a consensus sequence of a recombinase target site may, in some embodiments, be

the sequence most frequently bound, or bound with the highest affinity, by a given recombinase.

[0059] The term "engineered," as used herein refers to a protein molecule, a nucleic acid, complex, substance, or entity that has been designed, produced, prepared, synthesized, and/or manufactured by a human. Accordingly, an engineered product is a product that does not occur in nature.

[0060] The term "effective amount," as used herein, refers to an amount of a biologically active agent that is sufficient to elicit a desired biological response. In some embodiments, an effective amount of a recombinase may refer to the amount of the recombinase that is sufficient to induce recombination at a target site specifically bound and recombined by the recombinase. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.,* a nuclease, a recombinase, a hybrid protein, a fusion protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, the specific allele, genome, target site, cell, or tissue being targeted, and the agent being used.

[0061] A "guide nucleotide sequence-programmable DNA-binding protein," as used herein, refers to a protein, a polypeptide, or a domain that is able to bind DNA, and the binding to its target DNA sequence is mediated by a guide nucleotide sequence. The "guide nucleotide" may be an RNA or DNA molecule (*e.g.,* a single-stranded DNA or ssDNA molecule) that is complementary to the target sequence and can guide the DNA binding protein to the target sequence. As such, a guide nucleotide sequence-programmable DNA-binding protein may be a RNA-programmable DNA-binding protein, or an ssDNA-programmable DNA-binding protein. "Programmable" means the DNA-binding protein may be programmed to bind any DNA sequence that the guide nucleotide targets. The guide nucleotide sequence-programmable DNA-binding protein referred to herein may be any guide nucleotide sequence-programmable DNA-binding protein known in the art without limitation including, but not limited to, a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA-binding protein**.** The term "circularly permuted" refers to proteins in which the order of the amino acids in a protein has been altered, resulting in a protein structure with altered connectivity but a similar (overall) three-dimensional shape. Circular permutations are formed when the original n and c terminal amino acids are connected via a peptide bond; the peptide sequence is then broken in another location within the peptide sequence, causing a new n and c-terminus. Circular permutations may occur

through a number of processes including evolutionary events, post-translational modifications, or artificially engineered mutations. For example, circular permutations may be used to improve the catalytic activity or thermostability of proteins. A circularly permuted guide nucleotide sequence-programmable DNA-binding protein may be used with any of the embodiments described herein. The term "bifurcated" typically refers to a monomeric protein that is split into two parts. Typically both parts are required for the function of the monomeric protein. Bifurcated proteins may or may not dimerize on their own to reconstitute a functional protein. Bifurcations may occur through a number of processes including evolutionary events, post-translational modifications, or artificially engineered mutations. Other protein domains, when fused to bifurcated domains, can be used to force the reassembly of the bifurcated protein. In some cases, protein domains, whose interaction depends on a small molecule, can be fused to each bifurcated domain, resulting in the small-molecule regulated dimerization of the bifurcated protein.

[0062] The term "homologous," as used herein, is an art-understood term that refers to nucleic acids or polypeptides that are highly related at the level of nucleotide and/or amino acid sequence. Nucleic acids or polypeptides that are homologous to each other are termed "homologues." Homology between two sequences can be determined by sequence alignment methods known to those of skill in the art. In accordance with the invention, two sequences are considered to be homologous if they are at least about 50-60% identical, *e.g.,* share identical residues (*e.g.,* amino acid residues) in at least about 50-60% of all residues comprised in one or the other sequence, at least about 70% identical, at least about 80% identical, at least about 85% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical, for at least one stretch of at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 120, at least 150, or at least 200 amino acids.

[0063] The term "sequence identity" or "percent sequence identity" as used herein, may refer to the percentage of nucleic acid or amino acid residues within a given DNA or protein, respectively, that are identical to the reference sequence. See, for example: Christopher M. Holman, Protein Similarity Score: A Simplified Version of the BLAST Score as a Superior Alternative to Percent Identity for Claiming Genuses of Related Protein Sequences, 21 SANTA CLARA COMPUTER & HIGH TECH. L.J. 55, 60 (2004), which is herein incorporated by reference in its entirety.

[0064] The term "linker," as used herein, refers to a bond (*e.g.,* covalent bond), chemical group, or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a nuclease-inactive Cas9 domain and a nucleic acid-editing domain (*e.g.*, an adenosine deaminase). In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic-acid editing protein. In some embodiments, a linker joins a dCas9 and a nucleic-acid editing protein. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 7), which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGS (SEQ ID NO: 758). In some embodiments, a linker comprises $(SGGS)_n$ (SEQ ID NO: 758), $(GGGS)_n$ (SEQ ID NO: 759), $(GGGGS)_n$ (SEQ ID NO: 722), $(G)_n$, $(EAAAK)_n$ (SEQ ID NO: 723), $(GGS)_n$, or $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30, and wherein X is any amino acid. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15.

[0065] The term "mutation," as used herein, refers to a substitution of a residue within a sequence, *e.g.,* a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4^{th} ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[0066] The term "nuclear localization sequence" or "NLS" refers to an amino acid sequence that promotes import of a protein into the cell nucleus, for example, by nuclear transport. Nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, international PCT

application, PCT/EP2000/011690, filed November 23, 2000, published asWO/2001/038547 on May 31, 2001, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, a NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 702) or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 761).

[0067] The term "nuclease," as used herein, refers to an agent, for example, a protein, capable of cleaving a phosphodiester bond connecting two nucleotide residues in a nucleic acid molecule. In some embodiments, "nuclease" refers to a protein having an inactive DNA cleavage domain, such that the nuclease is incapable of cleaving a phosphodiester bond. In some embodiments, a nuclease is a protein, *e.g.,* an enzyme that can bind a nucleic acid molecule and cleave a phosphodiester bond connecting nucleotide residues within the nucleic acid molecule. A nuclease may be an endonuclease, cleaving a phosphodiester bonds within a polynucleotide chain, or an exonuclease, cleaving a phosphodiester bond at the end of the polynucleotide chain. In some embodiments, a nuclease is a site-specific nuclease, binding and/or cleaving a specific phosphodiester bond within a specific nucleotide sequence, which is also referred to herein as the "recognition sequence," the "nuclease target site," or the "target site." In some embodiments, a nuclease is a RNA-guided (*i.e.,* RNA-programmable) nuclease, which is associated with (*e.g.,* binds to) an RNA (*e.g.,* a guide RNA, "gRNA") having a sequence that complements a target site, thereby providing the sequence specificity of the nuclease. In some embodiments, a nuclease recognizes a single stranded target site, while in other embodiments, a nuclease recognizes a double-stranded target site, for example, a double-stranded DNA target site. The target sites of many naturally occurring nucleases, for example, many naturally occurring DNA restriction nucleases, are well known to those of skill in the art. A nuclease protein typically comprises a "binding domain" that mediates the interaction of the protein with the nucleic acid substrate, and also, in some cases, specifically binds to a target site, and a "cleavage domain" that catalyzes the cleavage of the phosphodiester bond within the nucleic acid backbone. In some embodiments a nuclease protein can bind and cleave a nucleic acid molecule in a monomeric form, while, in other embodiments, a nuclease protein has to dimerize or multimerize in order to cleave a target nucleic acid molecule. Binding domains and cleavage domains of naturally occurring nucleases, as well as modular binding domains and cleavage domains that can be fused to create nucleases binding specific target sites, are well known to those of skill in the art. For example, the binding domain of a guide nucleotide sequence-programmable DNA binding protein such as an RNA-programmable nucleases (*e.g.,* Cas9), or a Cas9 protein having an

inactive DNA cleavage domain, can be used as a binding domain (*e.g.*, that binds a gRNA to direct binding to a target site) to specifically bind a desired target site, and fused or conjugated to a cleavage domain.

[0068] The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.*, nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, gRNA, plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *i.e.*, analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5′ to 3′ direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.*, adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated

bases); intercalated bases; modified sugars (*e.g.*, 2′-fluororibose, ribose, 2′-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5′-*N*-phosphoramidite linkages).

[0069] The term "orthogonal" refers to biological components that interact minimally, if at all. Recombinase target sites containing different gRNA binding sites are orthogonal if the gRNA-directed recCas9 proteins do not interact, or interact minimally, with other potential recombinase sites. The term "orthogonality" refers to the idea that system components can be varied independently without affecting the performance of the other components. The gRNA directed nature of the complex makes the set of gRNA molecules complexed to recCas9 proteins capable of directing recombinase activity at only the gRNA-directed site. Orthogonality of the system is demonstrated by the complete or near complete dependence of the set of gRNA molecules on the enzymatic activity on a targeted recombinase site.

[0070] The term "pharmaceutical composition," as used herein, refers to a composition that can be administrated to a subject in the context of treatment and/or prevention of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, *e.g.*, a recombinase fused to a Cas9 protein, or fragment thereof (or a nucleic acid encoding a such a fusion), and optionally a pharmaceutically acceptable excipient. In some embodiments, a pharmaceutical composition comprises inventive Cas9 variant/fusion (*e.g.*, fCas9) protein(s) and gRNA(s) suitable for targeting the Cas9 variant/fusion protein(s) to a target nucleic acid. In some embodiments, the target nucleic acid is a gene. In some embodiments, the target nucleic acid is an allele associated with a disease, wherein the allele is cleaved by the action of the Cas9 variant/fusion protein(s). In some embodiments, the allele is an allele of the *CLTA* gene, the *VEGF* gene,the *PCDH15*, gene or the *FAM19A2* gene. See, *e.g.*, the Examples.

[0071] The term "proliferative disease," as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasms. Malignant neoplasia is also referred to as cancer. In some embodiments, the compositions and methods provided herein are useful for treating a proliferative disease. For example, in some embodiments, pharmaceutical compositions comprising Cas9 (*e.g.*, fCas9) protein(s) and gRNA(s) suitable for targeting the Cas9

protein(s) to an *VEGF* allele, wherein the allele is inactivated by the action of the Cas9 protein(s). See, *e.g.,* the Examples.

[0072] The terms "protein," "peptide," and "polypeptide" are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof. The term "fusion protein" as used herein refers to a hybrid polypeptide that comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an "amino-terminal fusion protein" or a "carboxy-terminal fusion protein," respectively. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference. A specific fusion protein referred to herein is recCas9, an RNA programmed small serine recombinase capable of functioning in mammalian cells created by fusion a catalytically inactive dCas9 to the catalytic domain of recombinase.

[0073] A "pseudo-gix" site or a "gix pseudo-site" as discussed herein is a specific pseudo-palindromic core DNA sequence that resembles the Gix recombinases' natural DNA recognition sequence. See, for example, N. D. F. Grindley, K. L. Whiteson, P. A. Rice, Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**, 567-605 (2006), which is incorporated by reference herein in its entirety. Similarly, a "pseudo-hix" or "hix-pseudo-site;" a "pseudo-six" or "six-pseudo site;" a "pseudo-resH" or "resH-pseudo-site;" "pseudo-

res" or "res-pseudo-site;" "pseudo-LoxP" or "LoxP-pseudo-site;" "pseudo-att" or "att-pseudo-site;" "pseudo-FTR" or "FTR-pseudo-site" is a specific pseudo-palindromic core DNA sequence that resembles the Hin recombinase's, β recombinase's, Sin recombinase's, Tn3 or γδ recombinase's, Cre recombinase's, λ phage integrase's, or FLP recombinase's natural DNA recognition sequence.

[0074] The terms "RNA-programmable nuclease" and "RNA-guided nuclease" are used interchangeably herein and refer to a nuclease that forms a complex with (*e.g.,* binds or associates with) one or more RNA that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though "gRNA" is used interchangeabley to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (*e.g.,* and directs binding of a Cas9 complex to the target); and (2) a domain that binds a Cas9 protein. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA, and comprises a stem-loop structure. For example, in some embodiments, domain (2) is homologous to a tracrRNA as depicted in Figure 1E of Jinek *et al.*, *Science* 337:816-821(2012), the entire contents of which is incorporated herein by reference. Other examples of gRNAs (*e.g.,* those including domain 2) can be found in U.S. Provisional Patent Application, U.S.S.N. 61/874,682, filed September 6, 2013, entitled "Switchable Cas9 Nucleases And Uses Thereof;" U.S. Provisional Patent Application, U.S.S.N. 61/874,746, filed September 6, 2013, entitled "Delivery System For Functional Nucleases;" PCT Application WO 2013/176722, filed March 15, 2013, entitled "Methods and Compositions for RNA-Directed Target DNA Modification and for RNA-Directed Modulation of Transcription;" and PCT Application WO 2013/142578, filed March 20, 2013, entitled "RNA-Directed DNA Cleavage by the Cas9-crRNA Complex;" the entire contents of each are hereby incorporated by reference in their entirety. Still other examples of gRNAs are provided herein. See *e.g.,* the Examples. In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an "extended gRNA." For example, an extended gRNA will *e.g.,* bind two or more Cas9 proteins and bind a target nucleic acid at two or more distinct regions, as described herein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the

nuclease/RNA complex to said target site, providing the sequence specificity of the nuclease:RNA complex. In some embodiments, the guide nucleotide sequence-programmable DNA binding protein is an RNA-programmable nuclease such as the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csn1) from *Streptococcus pyogenes* (see, *e.g.,* "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference.

[0075] Because RNA-programmable nucleases (*e.g.,* Cas9) use RNA:DNA hybridization to determine target DNA cleavage sites, these proteins are able to cleave, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (*e.g.,* to modify a genome) are known in the art (see *e.g.,* Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013); Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013); Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31**, 227-229 (2013); Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013); Dicarlo, J.E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. *et al.* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

[0076] The term "recombinase," as used herein, refers to a site-specific enzyme that mediates the recombination of DNA between recombinase recognition sequences, which results in the excision, integration, inversion, or exchange (*e.g.,* translocation) of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (*e.g.,* resolvases and invertases) and tyrosine recombinases (*e.g.,* integrases). Examples of serine recombinases include, without limitation,

Hin, Gin, Tn3, β-six, CinH, ParA, γδ, Bxb1, ϕC31, TP901, TG1, φBT1, R4, φRV1, φFC1, MR11, A118, U153, and gp29. Examples of tyrosine recombinases include, without limitation, Cre, FLP, R, Lambda, HK101, HK022, and pSAM2. The Gin recombinase referred to herein may be any Gin recombinase known in the art including, but not limited to, the Gin recombinases presented in T. Gaj *et al.*, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Research* **41**, 3937-3946 (2013), incorporated herein by reference in its entirety. In certain embodiments, the Gin recombinase catalytic domain has greater than 85%, 90%, 95%, 98%, or 99% sequence identity with the amino acid sequence shown in SEQ ID NO: 713. In another embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises a mutation corresponding to H106Y, and/or I127L, and/or I136R and/or G137F. In yet another embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises a mutation corresponding to H106Y, I127L, I136R, and G137F. In a further embodiment, the amino acid sequence of the Gin recombinase has been further mutated. In a specific embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises SEQ ID NO: 713. Gin recombinases bind to gix target sites (also referred to herein as "gix core," "minimal gix core," or "gix-related core" sequences). The minimal gix core recombinase site is NNNNAAASSWWSSTTTNNNN (SEQ ID NO: 19), wherein N is defined as any amino acid, W is an A or a T, and S is a G or a C. The gix target site may include any other mutations known in the art. In certain embodiments, the gix target site has greater than 90%, 95%, or 99% sequence identity with the amino acid sequence shown in SEQ ID NO: 19. The distance between the gix core or gix-related core sequence and at least one gRNA binding site may be from 1 to 10 base pairs, from 3 to 7 base pairs, from 5 to 7 base pairs, or from 5 to 6 base pairs. The distance between the gix core or gix-related core sequence and at least one gRNA binding site may be 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 base pairs.

[0077] The serine and tyrosine recombinase names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange. Recombinases have numerous applications, including the creation of gene knockouts/knock-ins and gene therapy applications. See, *e.g.*, Brown *et al.*, "Serine recombinases as tools for genome engineering." *Methods*. 2011;53(4):372-9; Hirano *et al.*, "Site-specific recombinases as tools for heterologous gene integration." *Appl. Microbiol. Biotechnol.* 2011; 92(2):227-39; Chavez and Calos, "Therapeutic applications of the ΦC31 integrase system." *Curr. Gene Ther.* 2011;11(5):375-81; Turan and Bode, "Site-specific recombinases: from tag-and-target- to

tag-and-exchange-based genomic modifications." *FASEB J.* 2011; 25(12):4088-107; Venken and Bellen, "Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and ΦC31 integrase." *Methods Mol. Biol.* 2012; 859:203-28; Murphy, "Phage recombinases and their applications." *Adv. Virus Res.* 2012; 83:367-414; Zhang *et al.*, "Conditional gene manipulation: Creating a new biological era." *J. Zhejiang Univ. Sci. B.* 2012; 13(7):511-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." *DNA Repair (Amst).* 2012; 1;11(10):781-8; the entire contents of each are hereby incorporated by reference in their entirety. The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in embodiments of the invention. The methods and compositions of the invention can be expanded by mining databases for new orthogonal recombinases or designing synthetic recombinases with defined DNA specificities (See, *e.g.*, Groth *et al.*, "Phage integrases: biology and applications." *J. Mol. Biol.* 2004; 335, 667-678; Gordley *et al.*, "Synthesis of programmable integrases." *Proc. Natl. Acad. Sci. USA.* 2009; 106, 5053-5058; the entire contents of each are hereby incorporated by reference in their entirety).

[0078] Other examples of recombinases that are useful in the methods and compositions described herein are known to those of skill in the art, and any new recombinase that is discovered or generated is expected to be able to be used in the different embodiments of the invention. In some embodiments, the catalytic domains of a recombinase are fused to a nuclease-inactivated RNA-programmable nuclease (*e.g.*, dCas9, or a fragment thereof), such that the recombinase domain does not comprise a nucleic acid binding domain or is unable to bind to a target nucleic acid that subsequently results in enzymatic catalysis (*e.g.*, the recombinase domain is engineered such that it does not have specific DNA binding activity). Recombinases lacking part of their DNA binding activity and those that act independently of accessory proteins and methods for engineering such are known, and include those described by Klippel *et al.*, "Isolation and characterisation of unusual gin mutants." *EMBO J.* 1988; 7: 3983–3989: Burke *et al.*, "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol Microbiol.* 2004; 51: 937–948; Olorunniji *et al.*, "Synapsis and catalysis by activated Tn3 resolvase mutants." *Nucleic Acids Res.* 2008; 36: 7181–7191; Rowland *et al.*, "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." *Mol Microbiol.* 2009; 74: 282–298; Akopian *et al.*, "Chimeric recombinases with designed DNA sequence recognition." *Proc Natl Acad Sci USA.* 2003;100: 8688–8691; Gordley *et al.*, "Evolution of programmable zinc finger-recombinases with activity in human

cells. *J Mol Biol.* 2007; 367: 802–813; Gordley *et al.*, "Synthesis of programmable integrases." *Proc Natl Acad Sci USA.* 2009;106: 5053–5058; Arnold *et al.*, "Mutants of Tn*3* resolvase which do not require accessory binding sites for recombination activity." *EMBO J.* 1999;18: 1407–1414; Gaj *et al.*, "Structure-guided reprogramming of serine recombinase DNA sequence specificity." *Proc Natl Acad Sci USA.* 2011;108(2):498-503; and Proudfoot *et al.*, "Zinc finger recombinases with adaptable DNA sequence specificity." *PLoS One.* 2011;6(4):e19537; the entire contents of each are hereby incorporated by reference. For example, serine recombinases of the resolvase-invertase group, *e.g.,* Tn3 and γδ resolvases and the Hin and Gin invertases, have modular structures with partly autonomous catalytic and DNA-binding domains (See, *e.g.,* Grindley *et al.*, "Mechanism of site-specific recombination." *Ann Rev Biochem.* 2006; 75: 567–605, the entire contents of which are incorporated by reference). The catalytic domains of these recombinases are therefore amenable to being recombined with nuclease-inactivated RNA-programmable nucleases (*e.g.,* dCas9, or a fragment thereof) as described herein, *e.g.,* following the isolation of 'activated' recombinase mutants which do not require any accessory factors (*e.g.,* DNA binding activities) (See, *e.g.,* Klippel *et al.*, "Isolation and characterisation of unusual gin mutants." *EMBO J.* 1988; 7: 3983–3989: Burke *et al.*, "Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol Microbiol.* 2004; 51: 937–948; Olorunniji *et al.*, "Synapsis and catalysis by activated Tn3 resolvase mutants." *Nucleic Acids Res.* 2008; 36: 7181–7191; Rowland *et al.*, "Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome." *Mol Microbiol.* 2009; 74: 282–298; Akopian *et al.*, "Chimeric recombinases with designed DNA sequence recognition." *Proc Natl Acad Sci USA.* 2003;100: 8688–8691).

[0079] Additionally, many other natural serine recombinases having an N-terminal catalytic domain and a C-terminal DNA binding domain are known (*e.g.,* phiC31 integrase, TnpX transposase, IS607 transposase), and their catalytic domains can be co-opted to engineer programmable site-specific recombinases as described herein (See, *e.g.,* Smith *et al.*, "Diversity in the serine recombinases." *Mol Microbiol.* 2002;44: 299–307, the entire contents of which are incorporated by reference). Similarly, the core catalytic domains of tyrosine recombinases (*e.g.,* Cre, λ integrase) are known, and can be similarly co-opted to engineer programmable site-specific recombinases as described herein (See, *e.g.,* Guo *et al.*, "Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse." *Nature.* 1997; 389:40–46; Hartung *et al.*, "Cre mutants with altered DNA binding properties." *J Biol Chem* 1998; 273:22884–22891; Shaikh *et al.*, "Chimeras of the Flp and

Cre recombinases: Tests of the mode of cleavage by Flp and Cre. *J Mol Biol.* 2000; 302:27–48; Rongrong *et al.*, "Effect of deletion mutation on the recombination activity of Cre recombinase." *Acta Biochim Pol.* 2005; 52:541–544; Kilbride *et al.*, "Determinants of product topology in a hybrid Cre-Tn3 resolvase site-specific recombination system." *J Mol Biol.* 2006; 355:185–195; Warren *et al.*, "A chimeric cre recombinase with regulated directionality." *Proc Natl Acad Sci USA.* 2008 105:18278–18283; Van Duyne, "Teaching Cre to follow directions." *Proc Natl Acad Sci USA.* 2009 Jan 6;106(1):4-5; Numrych *et al.*, "A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophage λ." *Nucleic Acids Res.* 1990; 18:3953–3959; Tirumalai *et al.*, "The recognition of core-type DNA sites by λ integrase." *J Mol Biol.* 1998; 279:513–527; Aihara *et al.*, "A conformational switch controls the DNA cleavage activity of λ integrase." *Mol Cell.* 2003; 12:187–198; Biswas *et al.*, "A structural basis for allosteric control of DNA recombination by λ integrase." *Nature.* 2005; 435:1059–1066; and Warren *et al.*, "Mutations in the amino-terminal domain of λ-integrase have differential effects on integrative and excisive recombination." *Mol Microbiol.* 2005; 55:1104–1112; the entire contents of each are incorporated by reference).

[0080] The term "recombine" or "recombination," in the context of a nucleic acid modification (*e.g.,* a genomic modification), is used to refer to the process by which two or more nucleic acid molecules, or two or more regions of a single nucleic acid molecule, are modified by the action of a recombinase protein (*e.g.,* an inventive recombinase fusion protein provided herein). Recombination can result in, *inter alia*, the insertion, inversion, excision, or translocation of nucleic acids, *e.g.,* in or between one or more nucleic acid molecules.

[0081] The term "recombinant" as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

[0082] The term "subject," as used herein, refers to an individual organism, for example, an individual mammal. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a

vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a research animal. In some embodiments, the subject is genetically engineered, *e.g.*, a genetically engineered non-human subject. The subject may be of either sex and at any stage of development. In some embodiments, the subject is genetically engineered, *e.g.*, a genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

[0083] The terms "target nucleic acid," and "target genome," as used herein in the context of nucleases, refer to a nucleic acid molecule or a genome, respectively, that comprises at least one target site of a given nuclease. In the context of fusions comprising a (nuclease-inactivated) RNA-programmable nuclease and a recombinase domain, a "target nucleic acid" and a "target genome" refers to one or more nucleic acid molecule(s), or a genome, respectively, that comprises at least one target site. In some embodiments, the target nucleic acid(s) comprises at least two, at least three, at least four, at least five, at least six, at least seven, or at least eight target sites. In some embodiments, the target nucleic acid(s) comprise four target sites.

[0084] The term "target site" refers to a sequence within a nucleic acid molecule that is bound and recombined (*e.g.*, at or nearby the target site) by a recombinase (*e.g.*, a dCas9-recombinase fusion protein provided herein). A target site may be single-stranded or double-stranded. For example, in some embodiments, four recombinase monomers are coordinated to recombine a target nucleic acid(s), each monomer being fused to a (nuclease-inactivated) Cas9 protein guided by a gRNA. In such an example, each Cas9 domain is guided by a distinct gRNA to bind a target nucleic acid(s), thus the target nucleic acid comprises four target sites, each site targeted by a separate dCas9-recombinase fusion (thereby coordinating four recombinase monomers which recombine the target nucleic acid(s)). For the RNA-guided nuclease-inactivated Cas9 (or gRNA-binding domain thereof) and inventive fusions of Cas9, the target site may be, in some embodiments, 17-20 base pairs plus a 3 base pair PAM (*e.g.*, NNN, wherein N independently represents any nucleotide). Typically, the first nucleotide of a PAM can be any nucleotide, while the two downstream nucleotides are specified depending on the specific RNA-guided nuclease. Exemplary target sites (*e.g.*, comprising a PAM) for RNA-guided nucleases, such as Cas9, are known to those of skill in the art and include, without limitation, NNG, NGN, NAG, and NGG, wherein each N is independently any nucleotide. In addition, Cas9 nucleases from different species (*e.g.*, *S. thermophilus* instead of *S. pyogenes*) recognize a PAM that comprises the sequence NGGNG (SEQ ID NO: 763). Additional PAM sequences are known, including, but not limited to,

NNAGAAW (SEQ ID NO: 749) and NAAR (SEQ ID NO: 771) (see, *e.g.,* Esvelt and Wang, Molecular Systems Biology, 9:641 (2013), the entire contents of which are incorporated herein by reference). In some aspects, the target site of an RNA-guided nuclease, such as, *e.g.,* Cas9, may comprise the structure [N_Z]-[PAM], where each N is independently any nucleotide, and z is an integer between 1 and 50, inclusive. In some embodiments, z is at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50. In some embodiments, z is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50. In some embodiments, z is 20. In certain embodiments, a "PAMless" RNA-guided nuclease (*e.g.*, a Pamless Cas9) or an RNA-guided nuclease with relaxed PAM requirements as further described herein may be used. In some embodiments, "target site" may also refer to a sequence within a nucleic acid molecule that is bound but not cleaved by a nuclease. For example, certain embodiments described herein provide proteins comprising an inactive (or inactivated) Cas9 DNA cleavage domain. Such proteins (*e.g.,* when also including a Cas9 RNA binding domain) are able to bind the target site specified by the gRNA; however, because the DNA cleavage site is inactivated, the target site is not cleaved by the particular protein. In some embodiments, such proteins are conjugated, fused, or bound to a recombinase (or a catalytic domain of a recombinase), which mediates recombination of the target nucleic acid. In some embodiments, the sequence actually cleaved or recombined will depend on the protein (*e.g.,* recombinase) or molecule that mediates cleavage or recombination of the nucleic acid molecule, and in some cases, for example, will relate to the proximity or distance from which the inactivated Cas9 protein(s) is/are bound.

[0085] The term "Transcriptional Activator-Like Effector," (TALE) as used herein, refers to bacterial proteins comprising a DNA binding domain, which contains a highly conserved 33-34 amino acid sequence comprising a highly variable two-amino acid motif (Repeat Variable Diresidue, RVD). The RVD motif determines binding specificity to a nucleic acid sequence and can be engineered according to methods known to those of skill in the art to specifically bind a desired DNA sequence (see, *e.g.,* Miller, Jeffrey; *et.al.* (February 2011). "A TALE nuclease architecture for efficient genome editing". *Nature Biotechnology* **29** (2): 143–8; Zhang, Feng; *et.al.* (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription" *Nature*

Biotechnology **29** (2): 149–53; Geiβler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". *PLoS ONE* **6** (5): e19509; Boch, Jens (February 2011). "TALEs of genome targeting". *Nature Biotechnology* **29** (2): 135–6; Boch, Jens; et.al. (December 2009). "Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors". *Science* **326** (5959): 1509–12; and Moscou, Matthew J.; Adam J. Bogdanove (December 2009). "A Simple Cipher Governs DNA Recognition by TAL Effectors" *Science* **326** (5959): 1501; the entire contents of each of which are incorporated herein by reference). The simple relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA binding domains by selecting a combination of repeat segments containing the appropriate RVDs.

[0086] The term "Transcriptional Activator-Like Element Nuclease," (TALEN) as used herein, refers to an artificial nuclease comprising a transcriptional activator-like effector DNA binding domain to a DNA cleavage domain, for example, a FokI domain. A number of modular assembly schemes for generating engineered TALE constructs have been reported (see *e.g.,* Zhang, Feng; *et.al.* (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". *Nature Biotechnology* **29** (2): 149–53; Geiβler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". *PLoS ONE* **6** (5): e19509; Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. *et al.* (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting". *Nucleic Acids Research*; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). "Assembly of custom TALE-type DNA binding domains by modular cloning". *Nucleic Acids Research*; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes". *Nucleic Acids Research.*; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). Bendahmane, Mohammed. ed. "Assembly of Designer TAL Effectors by Golden Gate Cloning". *PLoS ONE* **6** (5): e19722; the entire contents of each of which are incorporated herein by reference).

[0087] The terms "treatment," "treat," and "treating," refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms "treatment,"

"treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, *e.g.,* to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (*e.g.,* in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

[0088] The term "vector" refers to a polynucleotide comprising one or more recombinant polynucleotides of the present invention, *e.g.,* those encoding a Cas9 protein (or fusion thereof) and/or gRNA provided herein. Vectors include, but are not limited to, plasmids, viral vectors, cosmids, artificial chromosomes, and phagemids. The vector may be able to replicate in a host cell and may further be characterized by one or more endonuclease restriction sites at which the vector may be cut and into which a desired nucleic acid sequence may be inserted. Vectors may contain one or more marker sequences suitable for use in the identification and/or selection of cells which have or have not been transformed or genomically modified with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics (*e.g.,* kanamycin, ampicillin) or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (*e.g.,* β-galactosidase, alkaline phosphatase, or luciferase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies, or plaques. Any vector suitable for the transformation of a host cell (*e.g., E. coli*, mammalian cells such as CHO cell, insect cells, *etc.*) as embraced by the present invention, for example, vectors belonging to the pUC series, pGEM series, pET series, pBAD series, pTET series, or pGEX series. In some embodiments, the vector is suitable for transforming a host cell for recombinant protein production. Methods for selecting and engineering vectors and host cells for expressing proteins (*e.g.,* those provided herein), transforming cells, and expressing/purifying recombinant proteins are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4^{th} ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[0089] The term "zinc finger," as used herein, refers to a small nucleic acid-binding protein structural motif characterized by a fold and the coordination of one or more zinc ions that stabilize the fold. Zinc fingers encompass a wide variety of differing protein structures (see, *e.g.,* Klug A, Rhodes D (1987). "Zinc fingers: a novel protein fold for nucleic acid recognition". *Cold Spring Harb. Symp. Quant. Biol.* 52: 473–82, the entire contents of which are incorporated herein by reference). Zinc fingers can be designed to bind a specific sequence of nucleotides, and zinc finger arrays comprising fusions of a series of zinc fingers, can be designed to bind virtually any desired target sequence. Such zinc finger arrays can form a binding domain of a protein, for example, of a nuclease, *e.g.,* if conjugated to a nucleic acid cleavage domain. Different types of zinc finger motifs are known to those of skill in the art, including, but not limited to, Cys_2His_2, Gag knuckle, Treble clef, Zinc ribbon, Zn_2/Cys_6, and TAZ2 domain-like motifs (see, *e.g.,* Krishna SS, Majumdar I, Grishin NV (January 2003). "Structural classification of zinc fingers: survey and summary". *Nucleic Acids Res.* 31 (2): 532–50). Typically, a single zinc finger motif binds 3 or 4 nucleotides of a nucleic acid molecule. Accordingly, a zinc finger domain comprising 2 zinc finger motifs may bind 6-8 nucleotides, a zinc finger domain comprising 3 zinc finger motifs may bind 9-12 nucleotides, a zinc finger domain comprising 4 zinc finger motifs may bind 12-16 nucleotides, and so forth. Any suitable protein engineering technique can be employed to alter the DNA-binding specificity of zinc fingers and/or design novel zinc finger fusions to bind virtually any desired target sequence from 3 – 30 nucleotides in length (see, *e.g.,* Pabo CO, Peisach E, Grant RA (2001). "Design and selection of novel cys2His2 Zinc finger proteins". *Annual Review of Biochemistry* 70: 313–340; Jamieson AC, Miller JC, Pabo CO (2003). "Drug discovery with engineered zinc-finger proteins". *Nature Reviews Drug Discovery* 2 (5): 361–368; and Liu Q, Segal DJ, Ghiara JB, Barbas CF (May 1997). "Design of polydactyl zinc-finger proteins for unique addressing within complex genomes". *Proc. Natl. Acad. Sci. U.S.A.* 94 (11); the entire contents of each of which are incorporated herein by reference). Fusions between engineered zinc finger arrays and protein domains that cleave a nucleic acid can be used to generate a "zinc finger nuclease." A zinc finger nuclease typically comprises a zinc finger domain that binds a specific target site within a nucleic acid molecule, and a nucleic acid cleavage domain that cuts the nucleic acid molecule within or in proximity to the target site bound by the binding domain. Typical engineered zinc finger nucleases comprise a binding domain having between 3 and 6 individual zinc finger motifs and binding target sites ranging from 9 base pairs to 18 base pairs in length. Longer target

sites are particularly attractive in situations where it is desired to bind and cleave a target site that is unique in a given genome.

[0090] The term "zinc finger nuclease," as used herein, refers to a nuclease comprising a nucleic acid cleavage domain conjugated to a binding domain that comprises a zinc finger array. In some embodiments, the cleavage domain is the cleavage domain of the type II restriction endonuclease FokI. Zinc finger nucleases can be designed to target virtually any desired sequence in a given nucleic acid molecule for cleavage, and the possibility to design zinc finger binding domains to bind unique sites in the context of complex genomes allows for targeted cleavage of a single genomic site in living cells, for example, to achieve a targeted genomic alteration of therapeutic value. Targeting a double-strand break to a desired genomic locus can be used to introduce frame-shift mutations into the coding sequence of a gene due to the error-prone nature of the non-homologous DNA repair pathway. Zinc finger nucleases can be generated to target a site of interest by methods well known to those of skill in the art. For example, zinc finger binding domains with a desired specificity can be designed by combining individual zinc finger motifs of known specificity. The structure of the zinc finger protein Zif268 bound to DNA has informed much of the work in this field and the concept of obtaining zinc fingers for each of the 64 possible base pair triplets and then mixing and matching these modular zinc fingers to design proteins with any desired sequence specificity has been described (Pavletich NP, Pabo CO (May 1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A". *Science* **252** (5007): 809–17, the entire contents of which are incorporated herein). In some embodiments, separate zinc fingers that each recognizes a 3 base pair DNA sequence are combined to generate 3-, 4-, 5-, or 6-finger arrays that recognize target sites ranging from 9 base pairs to 18 base pairs in length. In some embodiments, longer arrays are contemplated. In other embodiments, 2-finger modules recognizing 6-8 nucleotides are combined to generate 4-, 6-, or 8- zinc finger arrays. In some embodiments, bacterial or phage display is employed to develop a zinc finger domain that recognizes a desired nucleic acid sequence, for example, a desired nuclease target site of 3-30 bp in length. Zinc finger nucleases, in some embodiments, comprise a zinc finger binding domain and a cleavage domain fused or otherwise conjugated to each other via a linker, for example, a polypeptide linker. The length of the linker determines the distance of the cut from the nucleic acid sequence bound by the zinc finger domain. If a shorter linker is used, the cleavage domain will cut the nucleic acid closer to the bound nucleic acid sequence, while a longer linker will result in a greater distance between the cut and the bound nucleic acid sequence. In some embodiments, the

cleavage domain of a zinc finger nuclease has to dimerize in order to cut a bound nucleic acid. In some such embodiments, the dimer is a heterodimer of two monomers, each of which comprise a different zinc finger binding domain. For example, in some embodiments, the dimer may comprise one monomer comprising zinc finger domain A conjugated to a FokI cleavage domain, and one monomer comprising zinc finger domain B conjugated to a FokI cleavage domain. In this non-limiting example, zinc finger domain A binds a nucleic acid sequence on one side of the target site, zinc finger domain B binds a nucleic acid sequence on the other side of the target site, and the dimerize FokI domain cuts the nucleic acid in between the zinc finger domain binding sites.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

[0091] The function and advantage of these and other embodiments of the present invention will be more fully understood from the Examples below. The following Examples are intended to illustrate the benefits of the present invention and to describe particular embodiments, but are not intended to exemplify the full scope of the invention. Accordingly, it will be understood that the Examples are not meant to limit the scope of the invention.

Guide nucleotide sequence-programmable DNA binding protein

[0092] The fusion proteins and methods described herein may use any programmable DNA binding domain.

[0093] In some embodiments, the programmable DNA binding protein domain comprises the DNA binding domain of a zinc finger nuclease (ZFN) or a transcription activator-like effector domain (TALE). In some embodiments, the programmable DNA binding protein domain may be programmed by a guide nucleotide sequence and is thus referred as a "guide nucleotide sequence-programmable DNA binding-protein domain." In some embodiments, the guide nucleotide sequence-programmable DNA binding protein is a nuclease inactive Cas9, or dCas9. A dCas9, as used herein, encompasses a Cas9 that is completely inactive in its nuclease activity, or partially inactive in its nuclease activity (*e.g.*, a Cas9 nickase). Thus, in some embodiments, the guide nucleotide sequence-programmable DNA binding protein is a Cas9 nickase. In some embodiments, the guide nucleotide sequence-programmable DNA binding protein is a nuclease inactive Cpf1. In some embodiments, the guide nucleotide sequence-programmable DNA binding protein is a nuclease inactive Argonaute.

[0094] In some embodiments, the guide nucleotide sequence-programmable DNA binding protein is a dCas9 domain. In some embodiments, the guide nucleotide sequence-programmable DNA binding protein is a Cas9 nickase. In some embodiments, the dCas9 domain comprises an amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 3. In some embodiments, the dCas9 domain comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 domains provided herein, and comprises mutations corresponding to D10X (X is any amino acid except for D) and/or H840X (X is any amino acid except for H) in SEQ ID NO: 1. In some embodiments, the dCas9 domain comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 domains provided herein, and comprises mutations corresponding to D10A and/or H840A in SEQ ID NO: 1. In some embodiments, the Cas9 nickase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 domains provided herein, and comprises mutations corresponding to D10X (X is any amino acid except for D) in SEQ ID NO: 1 and a histidine at a position correspond to position 840 in SEQ ID NO: 1. In some embodiments, the Cas9 nickase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 domains provided herein, and comprises mutations corresponding to D10A in SEQ ID NO: 1 and a histidine at a position correspond to position 840 in SEQ ID NO: 1. In some embodiments, variants or homologues of dCas9 or Cas9 nickase (*e.g.*, variants of SEQ ID NO: 2 or SEQ ID NO: 3, respectively) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to SEQ ID NO: 2 or SEQ ID NO: 3, respectively, and comprises mutations corresponding to D10A and/or H840A in SEQ ID NO: 1. In some embodiments, variants of Cas9 (*e.g.*, variants of SEQ ID NO: 2) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 2,

by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more, provided that the dCas9 variants comprise mutations corresponding to D10A and/or H840A in SEQ ID NO: 1. In some embodiments, variants of Cas9 nickase (*e.g.*, variants of SEQ ID NO: 3) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 3, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more, provided that the dCas9 variants comprise mutations corresponding to D10A and comprises a histidine at a position corresponding to position 840 in SEQ ID NO: 1.

[0095] Additional suitable nuclease-inactive dCas9 domains will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, D10A/D839A/H840A/N863A mutant domains in SEQ ID NO: 1 (See, *e.g.*, Prashant *et al.*, Nature Biotechnology. 2013; 31(9): 833-838, which is incorporated herein by reference), or K603R (See, *e.g.*, Chavez *et al.*, Nature Methods 12, 326–328, 2015, which is incorporated herein by reference).

[0096] In some embodiments, the nucleobase editors described herein comprise a Cas9 domain with decreased electrostatic interactions between the Cas9 domain and a sugar-phosphate backbone of a DNA, as compared to a wild-type Cas9 domain. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and a sugar-phosphate backbone of a DNA. In some embodiments, the nucleobase editors described herein comprises a dCas9 (*e.g.*, with D10A and H840A mutations in SEQ ID NO: 1) or a Cas9 nickase (*e.g.*, with D10A mutation in SEQ ID NO: 1), wherein the dCas9 or the Cas9 nickase further comprises one or more of a N497X, a R661X, a Q695X, and/or a Q926X mutation of the amino acid sequence provided in SEQ ID NO: 10, or a corresponding mutation in any of the amino acid sequences provided in SEQ ID NOs: 11-260, wherein X is any amino acid. In some embodiments, the nucleobase editors described herein comprises a dCas9 (*e.g.*, with D10A and H840A mutations in SEQ ID NO: 1) or a Cas9 nickase (*e.g.*, with D10A mutation in SEQ ID NO: 1), wherein the dCas9 or the Cas9 nickase further comprises one or more of a N497A, a R661A, a Q695A, and/or a Q926A mutation of the amino acid sequence provided in SEQ ID NO: 10, or a corresponding

mutation in any of the amino acid sequences provided in SEQ ID NOs: 11-260. In some embodiments, the Cas9 domain (*e.g.*, of any of the nucleobase editors provided herein) comprises the amino acid sequence as set forth in SEQ ID NO: 720. In some embodiments, the nucleobase editor comprises the amino acid sequence as set forth in SEQ ID NO: 721. Cas9 domains with high fidelity are known in the art and would be apparent to the skilled artisan. For example, Cas9 domains with high fidelity have been described in Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); and Slaymaker, I.M., *et al.* "Rationally engineered Cas9 nucleases with improved specificity." *Science* 351, 84-88 (2015); the entire contents of each are incorporated herein by reference.

Cas9 variant with decreased electrostatic interactions between the Cas9 and DNA backbone

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK
NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV
AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ

HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 720)

High fidelity nucleobase editor

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSET PGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFL KSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDN KVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS ELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLII KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 721)

[0097] The Cas9 protein recognizes a short motif (PAM motif) within the target DNA sequence, which is required for the Cas9-DNA interaction but that is not determined by complementarity to the guide RNA nucleotide sequence. A "PAM motif" or "protospacer adjacent motif," as used herein, refers to a DNA sequence adjacent to the 5'- or 3'- immediately following the DNA sequence that is complementary to the guide RNA oligonucleotide sequence. Cas9 will not successfully bind to, cleave, or nick the target DNA sequence if it is not followed by an appropriate PAM sequence. Without wishing to be bound by any particular theory, specific amino acid residues in the Cas9 enzyme are responsible for interacting with the bases of the PAM and determine the PAM specificity. Therefore, changes in these residues or nearby residues leads to a different or relaxed PAM specificity. Changing or relaxing the PAM specificity may shift the places where Cas9 can bind, as will be apparent to those of skill in the art based on the instant disclosure.

[0098] Wild-type Streptococcus pyogenes Cas9 recognizes a canonical PAM sequence (5′-NGG-3′). Other Cas9 nucleases (*e.g.*, Cas9 from Streptococcus thermophiles, Staphylococcus aureus, Neisseria meningitidis, or Treponema denticolaor) and Cas9 variants thereof have been described in the art to have different, or more relaxed PAM requirements. For example, in Kleinstiver *et al.*, Nature 523, 481–485, 2015; Klenstiver *et al.*, Nature 529, 490–495, 2016; Ran *et al.*, Nature, Apr 9; 520(7546): 186–191, 2015; Kleinstiver *et al.*, Nat Biotechnol, 33(12):1293-1298, 2015; Hou *et al.*, Proc Natl Acad Sci U S A, 110(39):15644-9, 2014; Prykhozhij *et al.*, PLoS One, 10(3): e0119372, 2015; Zetsche *et al.*, Cell 163, 759–771, 2015; Gao *et al.*, Nature Biotechnology, doi:10.1038/nbt.3547, 2016; Want *et al.*, Nature 461, 754–761, 2009; Chavez *et al.*, doi: dx.doi dot org/10.1101/058974; Fagerlund *et al.*, Genome Biol. 2015; 16: 25, 2015; Zetsche *et al.*, Cell, 163, 759–771, 2015; and Swarts *et al.*, Nat Struct Mol Biol, 21(9):743-53, 2014, each of which is incorporated herein by reference.

[0099] Thus, the guide nucleotide sequence-programmable DNA-binding protein of the present disclosure may recognize a variety of PAM sequences including, without limitation PAM sequences that are on the 3' or the 5' end of the DNA sequence determined by the guide RNA. For example, the sequence may be: NGG, NGAN (SEQ ID NO: 741), NGNG (SEQ ID NO: 742), NGAG (SEQ ID NO: 743), NGCG (SEQ ID NO: 744), NNGRRT (SEQ ID NO: 745), NGRRN (SEQ ID NO: 746), NNNRRT (SEQ ID NO: 747), NNNGATT (SEQ ID NO: 748), NNAGAAW (SEQ ID NO: 749), NAAAC (SEQ ID NO:

750), TTN, TTTN (SEQ ID NO: 751), and YTN, wherein Y is a pyrimidine, R is a purine, and N is any nucleobase.

[00100] Some aspects of the disclosure provide RNA-programmable DNA binding proteins, which may be used to guide a protein, such as a base editor, to a specific nucleic acid (*e.g.,* DNA or RNA) sequence. Nucleic acid programmable DNA binding proteins include, without limitation, Cas9 (*e.g.,* dCas9 and nCas9), CasX, CasY, Cpf1, C2c1, C2c2, C2C3, and Argonaute. One example of an RNA-programmable DNA-binding protein that has different PAM specificity is Clustered Regularly Interspaced Short Palindromic Repeats from Prevotella and Francisella 1 (Cpf1). Similar to Cas9, Cpf1 is also a class 2 CRISPR effector. It has been shown that Cpf1 mediates robust DNA interference with features distinct from Cas9. Cpf1 is a single RNA-guided endonuclease lacking tracrRNA, and it may utilize a T-rich protospacer-adjacent motif (*e.g.*, TTN, TTTN (SEQ ID NO: 751), or YTN), which is on the 5'-end of the DNA sequence determined by the guide RNA. Moreover, Cpf1 cleaves DNA via a staggered DNA double-stranded break. Out of 16 Cpf1-family proteins, two enzymes from Acidaminococcus and Lachnospiraceae are shown to have efficient genome-editing activity in human cells. Cpf1 proteins are known in the art and have been described previously, for example Yamano *et al.*, "Crystal structure of Cpf1 in complex with guide RNA and target DNA." *Cell* (165) 2016, p. 949-962; the entire contents of which is hereby incorporated by reference.

[00101] Also useful in the present compositions and methods are nuclease-inactive Cpf1 (dCpf1) variants that may be used as a guide nucleotide sequence-programmable DNA-binding protein domain. The Cpf1 protein has a RuvC-like endonuclease domain that is similar to the RuvC domain of Cas9 but does not have a HNH endonuclease domain, and the N-terminal of Cpf1 does not have the alfa-helical recognition lobe of Cas9. It was shown in Zetsche *et al.*, Cell, 163, 759–771, 2015 (which is incorporated herein by reference) that, the RuvC-like domain of Cpf1 is responsible for cleaving both DNA strands and inactivation of the RuvC-like domain inactivates Cpf1 nuclease activity. For example, mutations corresponding to D917A, E1006A, or D1255A in Francisella novicida Cpf1 (SEQ ID NO: 714) inactivate Cpf1 nuclease activity. In some embodiments, the dCpf1 of the present disclosure may comprise mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/ E1006A/D1255A in SEQ ID NO: 714. In other embodiments, the Cpf1 nickase of the present disclosure may comprise mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/ E1006A/D1255A in SEQ ID NO: 714. A Cpf1 nickase useful

for the embodiments of the instant disclosure may comprise other mutations and/or further mutations known in the field. It is to be understood that any mutations, *e.g.*, substitution mutations, deletions, or insertions that fully or partially inactivates the RuvC domain of Cpf1 may be used in accordance with the present disclosure, and that these mutations of Cpf1 may result in, for example, a dCpf1 or Cpf1 nickase.

[00102] Thus, in some embodiments, the guide nucleotide sequence-programmable DNA binding protein is a nuclease inactive Cpf1 (dCpf1). In some embodiments, the dCpf1 comprises an amino acid sequence of any one SEQ ID NOs: 714-717. In some embodiments, the dCpf1 comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any one of SEQ ID NOs: 714-717, and comprises mutations corresponding to D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, or D917A/ E1006A/D1255A in SEQ ID NO: 714. Cpf1 from other bacterial species may also be used in accordance with the present disclosure, as a dCpf1 or Cpf1 nickase.

Wild type *Francisella novicida* Cpf1 (SEQ ID NO: 714) (D917, E1006, and D1255 are bolded and underlined)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDG

KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

Francisella novicida Cpf1 D917A (SEQ ID NO: 715)
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**A**RGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

Francisella novicida Cpf1 E1006A (SEQ ID NO: 716)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI
DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**D**ANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

Francisella novicida Cpf1 D1255A (SEQ ID NO: 717)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYH
QFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSE
KFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWT
TYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGEN
TKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTM
QSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDY
SVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDI

DKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIK
DLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYI
TQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFD
DKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKN
GSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVE
NQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDER
NLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSI**D**RGERHLAYYTLVDG
KGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQV
VHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEF
DKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESV
SKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKN
HNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQM
RNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA**A**ANGAYHIGLKGLMLLGRI
KNNQEGKKLNLVIKNEEYFEFVQNRNN

[00103] In addition to Cas9 and Cpf1, three distinct Class 2 CRISPR-Cas systems (C2c1, C2c2, and C2c3) have been described by Shmakov et al., "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", Mol. Cell, 2015 Nov 5; 60(3): 385–397, the entire contents of which is hereby incorporated by reference. Effectors of two of the systems, C2c1 and C2c3, contain RuvC-like endonuclease domains related to Cpf1. A third system, C2c2 contains an effector with two predicated HEPN RNase domains. Production of mature CRISPR RNA is tracrRNA-independent, unlike production of CRISPR RNA by C2c1. C2c1 depends on both CRISPR RNA and tracrRNA for DNA cleavage. Bacterial C2c2 has been shown to possess a unique RNase activity for CRISPR RNA maturation distinct from its RNA-activated single-stranded RNA degradation activity. These RNase functions are different from each other and from the CRISPR RNA-processing behavior of Cpf1. See, *e.g.*, East-Seletsky, et al., "Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection", *Nature*, 2016 Oct 13;538(7624):270-273, the entire contents of which are hereby incorporated by reference. In vitro biochemical analysis of C2c2 in *Leptotrichia shahii* has shown that C2c2 is guided by a single CRISPR RNA and can be programed to cleave ssRNA targets carrying complementary protospacers. Catalytic residues in the two conserved HEPN domains mediate cleavage. Mutations in the catalytic residues generate catalytically inactive RNA-binding proteins. See

e.g., Abudayyeh et al., "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector", *Science*, 2016 Aug 5; 353(6299), the entire contents of which are hereby incorporated by reference.

[00104] The crystal structure of *Alicyclobaccillus acidoterrastris* C2c1 (AacC2c1) has been reported in complex with a chimeric single-molecule guide RNA (sgRNA). See, *e.g.*, Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell*, 2017 Jan 19;65(2):310-322, the entire contents of which are hereby incorporated by reference. The crystal structure has also been reported in *Alicyclobacillus acidoterrestris* C2c1 bound to target DNAs as ternary complexes. See, *e.g.*, Yang et al., "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell*, 2016 Dec 15;167(7):1814-1828, the entire contents of which are hereby incorporated by reference. Catalytically competent conformations of AacC2c1, both with target and non-target DNA strands, have been captured independently positioned within a single RuvC catalytic pocket, with C2c1-mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between C2c1 ternary complexes and previously identified Cas9 and Cpf1 counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems.

[00105] In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein of any of the fusion proteins provided herein may be a C2c1, a C2c2, or a C2c3 protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein is a C2c1 protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein is a C2c2 protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein is a C2c3 protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring C2c1, C2c2, or C2c3 protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein is a naturally-occurring C2c1, C2c2, or C2c3 protein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any of the C2c1, C2c2, or C2c3 proteins described herein. In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein comprises an amino acid sequence

of any one of the C2c1, C2c2, or C2c3 proteins described herein. It should be appreciated that C2c1, C2c2, or C2c3 from other bacterial species may also be used in accordance with the present disclosure.

C2c1 (uniprot.org/uniprot/T0D7A2#)
sp|T0D7A2|C2C1_ALIAG CRISPR-associated endonuclease C2c1 OS=Alicyclobacillus acidoterrestris (strain ATCC 49025 / DSM 3922 / CIP 106132 / NCIMB 13137 / GD3B) GN=c2c1 PE=1 SV=1
MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNG
DGEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAI
GAKGDAQQIARKFLSPLADKDAVGGLGIAKAGNKPRWVRMREAGEPGWEEEKEKA
ETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDM
KEASPGLESKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNT
RRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMFATFTLPDAT
AHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISM
SEQLDNLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARD
VYLNVSVRVQSQSEARGERRPPYAAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKL
GSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAV
HERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSW
AKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRH
MGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIEYLERQYKFLKSWSFFGKVS
GQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALDERGKGKWVA
KYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTM
YAAFSSRFDARTGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRAD
DLIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWGEVD
GELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKLSEEEAELL
VEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQD
SACENTGDI (SEQ ID NO: 762)

C2c2 (uniprot.org/uniprot/P0DOC6)
>sp|P0DOC6|C2C2_LEPSD CRISPR-associated endoribonuclease C2c2 OS=Leptotrichia shahii (strain DSM 19757 / CCUG 47503 / CIP 107916 / JCM 16776 / LB37) GN=c2c2 PE=1 SV=1
MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKIDNNKFIR
KYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKS
EKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDCSIILRI
IENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEEHLREKLLKDDKIDVI
LTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIAD
FVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKI
VKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDS
KKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKILK
RVKQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFASTNMELNK
IFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTKI
GTNERNRILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNII
TKINDIKISEENNNDIKYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKE
LYKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIENYYKNAQISASKGNNKAIK
KYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTI
VINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNEC
ITENWNLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDV
LEKKLEKIVIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKS
KILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLFLNIG
NPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKE
KYIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLIDIN
WKLAIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYK
FFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFYIVRNPFADYSIAEQIDRV
SNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSV
LELESYNSDYIKNLIIELLTKIENTNDTL (SEQ ID NO: 764)

[00106] In some embodiments, the guide nucleotide sequence-programmable DNA-binding protein domain of the present disclosure has no requirements for a PAM sequence. One example of such a guide nucleotide sequence-programmable DNA-binding protein may be an Argonaute protein from Natronobacterium gregoryi (NgAgo). NgAgo is a ssDNA-guided endonuclease. NgAgo binds 5′ phosphorylated ssDNA of ~24 nucleotides (gDNA) to

guide it to its target site and will make DNA double-strand breaks at the gDNA site. In contrast to Cas9, the NgAgo–gDNA system does not require a protospacer-adjacent motif (PAM). Using a nuclease inactive NgAgo (dNgAgo) can greatly expand the codons that may be targeted. The characterization and use of NgAgo have been described in Gao *et al.*, Nat Biotechnol., 2016 Jul;34(7):768-73. PubMed PMID: 27136078; Swarts *et al.*, Nature. 507(7491) (2014):258-61; and Swarts *et al.*, Nucleic Acids Res. 43(10) (2015):5120-9, each of which is incorporated herein by reference. The sequence of Natronobacterium gregoryi Argonaute is provided in SEQ ID NO: 718.

Wild type *Natronobacterium gregoryi* Argonaute (SEQ ID NO: 718)
MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITL
WKNTTPKDVFTYDYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTD
EDETFAGGEPLDHHLDDALNETPDDAETESDSGHVMTSFASRDQLPEWTLHTYTLT
ATDGAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLDCG
VRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCD
EFDLHERYDLSVEVGHSGRAYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPR
RGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQTPINTDLLDAIEAADRRVVETRR
QGHGDDAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAFAER
LDPVRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVN
PPESFEVAVVLPEQQADTCKAQWDTMADLLNQAGAPPTRSETVQYDAFSSPESISLN
VAGAIDPSEVDAAFVVLPPDQEGFADLASPTETYDELKKALANMGIYSQMAYFDRF
RDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINI
AATATAVYKDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHR
DGFMNEDLDPATEFLNEQGVEYDIVEIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEP
RATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVYLLSQSHIQVHNSTA
RLPITTAYADQASTHATKGYLVQTGAFESNVGFL

[00107] Also provided herein are Cas9 variants that have relaxed PAM requirements (PAMless Cas9). PAMless Cas9 exhibits an increased activity on a target sequence that does not include a canonical PAM (*e.g.*, NGG) sequence at its 3′-end as compared to Streptococcus pyogenes Cas9 as provided by SEQ ID NO: 1, *e.g.*, increased activity by at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold. Such Cas9 variants that have relaxed PAM

requirements are described in US Provisional Applications, USSN 62/245,828, filed October 23, 2015; 62/279,346, filed January 15, 2016; 62/311,763, filed March 22, 2016; 62/322,178, filed April 13, 2016; and 62/357,332, filed June 30, 2016, each of which is incorporated herein by reference. In some embodiments, the dCas9 or Cas9 nickase useful in the present disclosure may further comprise mutations that relax the PAM requirements, *e.g.*, mutations that correspond to A262T, K294R, S409I, E480K, E543D, M694I, or E1219V in SEQ ID NO: 1.

[00108] The "-" used in the general architecture discussed herein may indicate the presence of an optional linker. The term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a guide nucleotide sequence-programmable DNA binding protein domain and a recombinase catalytic domain. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated. Linkers may be of any form known in the art. For example, the linker may be a linker from a website such as www[dot]ibi[dot]vu[dot]nl/programs/linkerdbwww/ or from www[dot] ibi[dot]vu[dot]nl/programs/linkerdbwww/src/database.txt. The linkers may also be unstructured, structured, helical, or extended.

[00109] In some embodiments, the guide nucleotide sequence-programmable DNA binding protein domain and the recombinase catalytic domain are fused to each other via a linker. Various linker lengths and flexibilities between the guide nucleotide sequence-programmable DNA binding protein domain and the recombinase catalytic domain can be employed (e.g., ranging from flexible linkers of the form (GGGS)n (SEQ ID NO: 759), (GGGGS)n (SEQ ID NO: 722), (GGS)n, and (G)n to more rigid linkers of the form (EAAAK)n (SEQ ID NO: 723), SGSETPGTSESATPES (SEQ ID NO: 724) (see, e.g., Guilinger *et al.*, Nat. Biotechnol. 2014; 32(6): 577-82; the entire contents of which is incorporated herein by reference), (XP)n, or a combination of any of these, wherein X is any amino acid, and n is independently an integer between 1 and 30, in order to achieve the optimal length for activity for the specific application. In some embodiments, n is

independently 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, or 30, or, if more than one linker or more than one linker motif is present, any combination thereof. In some embodiments, the linker comprises a (GGS)n motif, wherein n is 1,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15. In some embodiments, the linker comprises a (GGS)n motif, wherein n is 1, 3, or 7. In some embodiments, the linker comprises an XTEN linker. The XTEN linker may have the sequence SGSETPGTSESATPES (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9). In some embodiments, the linker comprises an amino acid sequence chosen from the group including, but not limited to, AGVF (SEQ ID NO: 772), GFLG (SEQ ID NO: 773), FK, AL, ALAL (SEQ ID NO: 774), and ALALA (SEQ ID NO: 775). In some embodiments, suitable linker motifs and configurations include those described in Chen *et al.*, Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013; 65(10):1357-69, which is incorporated herein by reference. In some embodiments, the linker may comprise any of the following amino acid sequences: VPFLLEPDNINGKTC (SEQ ID NO: 10), GSAGSAAGSGEF (SEQ ID NO: 11), SIVAQLSRPDPA (SEQ ID NO: 12), MKIIEQLPSA (SEQ ID NO: 13), VRHKLKRVGS (SEQ ID NO: 14), GHGTGSTGSGSS (SEQ ID NO: 15), MSRPDPA (SEQ ID NO: 16), GSAGSAAGSGEF (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9), and GGSM (SEQ ID NO: 17).

[00110] Additional suitable linker sequences will be apparent to those of skill in the art based on the instant disclosure. In certain embodiments, the linker may have a length of about 33 angstroms to about 81 angstroms. In another embodiment, the linker may have a length of about 54 angstroms to about 81 angstroms. In a further embodiment, the linker may have a length of about 63 to about 81 angstroms. In another embodiment, the linker may have a length of about 65 angstroms to about 75 angstroms. In some embodiments, the linker may have a weight of about 1.20 kDa to about 1.85 kDa. In certain embodiments, the linker may have a weight of about 1.40 kDa to about 1.85 kDa. In certain embodiments, the linker may have a weight of about 1.60 kDa to about 1.7 kDa. In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.,* a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is a peptide linker. In some embodiments, the peptide linker is any stretch of amino acids having at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, or more amino acids. In certain embodiments, the peptide linker is from

18 to 27 amino acids long. In a specific embodiment, the peptide linker is 24 amino acids long. In some embodiments, the peptide linker comprises repeats of the tri-peptide Gly-Gly-Ser, *e.g.,* comprising the sequence $(GGS)_n$, wherein n represents at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more repeats. In some embodiments, the linker comprises the sequence $(GGS)_6$ (SEQ ID NO: 6). In some embodiments, the peptide linker is the 16 residue "XTEN" linker, or a variant thereof (See, *e.g.,* the Examples; and Schellenberger *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* **27,** 1186–1190 (2009)). In some embodiments, the XTEN linker comprises the sequence SGSETPGTSESATPES (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9). In some embodiments, the peptide linker is selected from VPFLLEPDNINGKTC (SEQ ID NO: 10), GSAGSAAGSGEF (SEQ ID NO: 11), SIVAQLSRPDPA (SEQ ID NO: 12), MKIIEQLPSA (SEQ ID NO: 13), VRHKLKRVGS (SEQ ID NO: 14), GHGTGSTGSGSS (SEQ ID NO: 15), MSRPDPA (SEQ ID NO: 16); or GGSM (SEQ ID NO: 17). In some embodiments, the linker is a non-peptide linker. In certain embodiments, the non-peptide linker comprises one or more of polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacryl amide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker. In one embodiment, the alkyl linker has the formula —NH—$(CH_2)_s$—C(O)—, wherein s may be any integer. In a further embodiment, s may be any integer from 1-20.

Recombinase catalytic domain

[00111] The recombinase catalytic domain for use in the compositions and methods of the instant disclosure may be from any recombinase. Suitable recombinases catalytic domains for use in the disclosed methods and compositions may be obtained from, for example, and without limitation, tyrosine recombinases and serine recombinases. Some exemplary suitable recombinases provided herein include, for example, and without limitation, Gin recombinase (acting on gix sites), Hin recombinase (acting on hix sites), β recombinase (acting on six sites), Sin recombinase (acting on resH sites), Tn3 recombinase (acting on res sites), γδ recombinase (acting on res sites), Cre recombinase from bacteriophage P1 (acting on LoxP sites); FLP recombinases of fungal origin (acting on FTR sites); and phiC31 integrase (acting on att sites). Non-limiting sequences of exemplary suitable recombinases may be found below.

Cre recombinase sequence

MSNLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAA
WCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDS
NAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRNLAFLGI
AYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWI
SVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQR
YLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVR
LLEDGD (SEQ ID NO: 725)

FLP recombinase

MPQFGILCKTPPKVLVRQFVERFERPSGEKIALCAAELTYLCWMITHNGTAIKRATF
MSYNTIISNSLSFDIVNKSLQFKYKTQKATILEASLKKLIPAWEFTIIPYYGQKHQSDIT
DIVSSLQLQFESSEEADKGNSHSKKMLKALLSEGESIWEITEKILNSFEYTSRFTKTKT
LYQFLFLATFINCGRFSDIKNVDPKSFKLVQNKYLGVIIQCLVTETKTSVSRHIYFFSA
RGRIDPLVYLDEFLRNSEPVLKRVNRTGNSSSNKQEYQLLKDNLVRSYNKALKKNA
PYSIFAIKNGPKSHIGRHLMTSFLSMKGLTELTNVVGNWSDKRASAVARTTYTHQIT
AIPDHYFALVSRYYAYDPISKEMIALKDETNPIEEWQHIEQLKGSAEGSIRYPAWNGII
SQEVLDYLSSYINRRI (SEQ ID NO: 726)

γδ recombinase (Gamma Delta resolvase)

MRLFGYARVSTSQQSLDIQVRALKDAGVKANRIFTDKASGSSSDRKGLDLLRMKVE
EGDVILVKKLDRLGRDTADMIQLIKEFDAQGVSIRFIDDGISTDGEMGKMVVTILSAV
AQAERQRILERTNEGRQEAMAKGVVFGRKR (SEQ ID NO: 727)

γδ recombinase (E124Q mutation)

MRLFGYARVSTSQQSLDIQVRALKDAGVKANRIFTDKASGSSSDRKGLDLLRMKVE
EGDVILVKKLDRLGRDTADMIQLIKEFDAQGVSIRFIDDGISTDGEMGKMVVTILSAV
AQAERQRILQRTNEGRQEAMAKGVVFGRKR (SEQ ID NO: 728)

γδ recombinase (E102Y/E124Q mutation)

MRLFGYARVSTSQQSLDIQVRALKDAGVKANRIFTDKASGSSSDRKGLDLLRMKVE
EGDVILVKKLDRLGRDTADMIQLIKEFDAQGVSIRFIDDGISTDGYMGKMVVTILSAV
AQAERQRILQRTNEGRQEAMAKGVVFGRKR (SEQ ID NO: 729)

β recombinase
MAKIGYARVSSKEQNLDRQLQALQGVSKVFSDKLSGQSVERPQLQAMLNYIREGDI
VVVTELDRLGRNNKELTELMNAIQQKGATLEVLDLPSMNGIEDENLRRLINNLVIEL
YKYQAESERKRIKERQAQGIEIAKSKGKFKGRQH (SEQ ID NO: 730)

β recombinase (N95D
mutation)MAKIGYARVSSKEQNLDRQLQALQGVSKVFSDKLSGQSVERPQLQAMLNY
IREGDIVVVTELDRLGRNNKELTELMNAIQQKGATLEVLDLPSMDGIEDENLRRLINN
LVIELYKYQAESERKRIKERQAQGIEIAKSKGKFKGRQH (SEQ ID NO: 731)

Sin recombinase
MIIGYARVSSLDQNLERQLENLKTFGAEKIFTEKQSGKSIENRPILQKALNFVRMGDR
FIVESIDRLGRNYNEVIHTVNYLKDKEVQLMITSLPMMNEVIGNPLLDKFMKDLIIQIL
AMVSEQERNESKRRQAQGIQVAKEKGVYKGRPL (SEQ ID NO: 732)

Sin recombinase (Q87R/Q115R mutations)
MIIGYARVSSLDQNLERQLENLKTFGAEKIFTEKQSGKSIENRPILQKALNFVRMGDR
FIVESIDRLGRNYNEVIHTVNYLKDKEVRLMITSLPMMNEVIGNPLLDKFMKDLIIRIL
AMVSEQERNESKRRQAQGIQVAKEKGVYKGRPL (SEQ ID NO: 733)

Tn3 recombinase
MRLFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVK
EGDVILVKKLDRLGRDTADMLQLIKEFDAQGVAVRFIDDGISTDGDMGQMVVTILS
AVAQAERRRILERTNEGRQEAKLKGIKFGRRR (SEQ ID NO: 734)

Tn3 recombinase (G70S/D102Y, E124Q mutations)
MRLFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVK
EGDVILVKKLDRLSRDTADMLQLIKEFDAQGVAVRFIDDGISTDGYMGQMVVTILSA
VAQAERRRILQRTNEGRQEAKLKGIKFGRRR (SEQ ID NO: 735)

Hin recombinase
MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPGLKRALKYVNKGDT
LVVWKLDRLGRSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFHVMSALAE
MERELIVERTLAGLAAARAQGRLGGRPV (SEQ ID NO: 736)

Hin recombinase (H107Y mutation)
MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPGLKRALKYVNKGDT
LVVWKLDRLGRSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFYVMSALAE
MERELIVERTLAGLAAARAQGRLGGRPV (SEQ ID NO: 737)

PhiC31 recombinase
MDTYAGAYDRQSRERENSSAASPATQRSANEDKAADLQREVERDGGRFRFVGHFSE
APGTSAFGTAERPEFERILNECRAGRLNMIIVYDVSRFSRLKVMDAIPIVSELLALGVT
IVSTQEGVFRQGNVMDLIHLIMRLDASHKESSLKSAKILDTKNLQRELGGYVGGKAP
YGFELVSETKEITRNGRMVNVVINKLAHSTTPLTGPFEFEPDVIRWWWREIKTHKHL
PFKPGSQAAIHPGSITGLCKRMDADAVPTRGETIGKKTASSAWDPATVMRILRDPRIA
GFAAEVIYKKKPDGTPTTKIEGYRIQRDPITLRPVELDCGPIIEPAEWYELQAWLDGR
GRGKGLSRGQAILSAMDKLYCECGAVMTSKRGEESIKDSYRCRRRKVVDPSAPGQH
EGTCNVSMAALDKFVAERIFNKIRHAEGDEETLALLWEAARRFGKLTEAPEKSGERA
NLVAERADALNALEELYEDRAAGAYDGPVGRKHFRKQQAALTLRQQGAEERLAEL
EAAEAPKLPLDQWFPEDADADPTGPKSWWGRASVDDKRVFVGLFVDKIVVTKSTT
GRGQGTPIEKRASITWAKPPTDDDEDDAQDGTEDVAATGA (SEQ ID NO: 738)

[00112] Recombinases for use with the disclosed compositions and methods may also include further mutations. Some aspects of this disclosure provide recombinases comprising an amino acid sequence that is at least 70%, at least 80%, at least 90%, at least 95%, or at least 97% identical to the sequence of the recombinase sequence discussed herein, wherein the amino acid sequence of the recombinase comprises at least one mutation as compared to the sequence of the recombinase sequence discussed herein. In some embodiments, the amino acid sequence of the recombinase comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 mutations as compared to the sequence of the recombinase sequence discussed herein.

[00113] For example, the γδ recombinase may comprise one or more mutations from the list: R2A, E56K, G101S, E102Y, M103I, or E124Q. In one embodiment, the γδ recombinase may comprise an E102Y mutation, an E124Q mutation, or both an E102Y and E124Q mutation. In another embodiment, the β recombinase may comprise one or more mutations including, but not limited to N95D. See, for example, Sirk *et al.*, "Expanding the zinc-finger recombinase repertoire: directed evolution and mutational analysis of serine recombinase specificity determinants" Nucl Acids Res (2014) 42 (7): 4755-4766. In another embodiment, the Sin recombinase may have one or more mutations including, but not limited to: Q87R, Q115R, or Q87R and Q115R. In another embodiment, the Tn3 recombinase may have one or more mutations including, but not limited to: G70S, D102Y, E124Q, and any combination thereof. In another embodiment, the Hin recombinase may have one or more mutations including, but not limited to: H107Y. In another embodiment, the Sin recombinase may have one or more mutations including, but not limited to: H107Y.Any of the recombinase catalytic domains for use with the disclosed compositions and methods may have greater than 85%, 90%, 95%, 98%, or 99% sequence identity with the native (or wild type) amino acid sequence. For example, in certain embodiments, the Gin recombinase catalytic domain has greater than 85%, 90%, 95%, 98%, or 99% sequence identity with the amino acid sequence shown in SEQ ID NO: 713. In another embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises a mutation corresponding to H106Y, and/or I127L, and/or I136R and/or G137F. In yet another embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises a mutation corresponding to H106Y, I127L, I136R, and G137F. In a further embodiment, the amino acid sequence of the Gin recombinase has been further mutated. In a specific embodiment, the amino acid sequence of the Gin recombinase catalytic domain comprises SEQ ID NO: 713.

[00114] The recombinase catalytic domain for use in the compositions and methods of the instant disclosure may be from an evolved recombinase. As used herein, the term "evolved recombinase" refers to a recombinase that has been altered (*e.g.*, through mutation) to recognize non-native DNA target sequences.

[00115] Suitable recombinases that can be evolved include, for example, and without limitation, tyrosine recombinases and serine recombinases (*e.g.*, any of the recombinases discussed herein). Some exemplary suitable recombinases that can be evolved by the methods and strategies provided herein include, for example, and without limitation, Gin recombinase (acting on gix sites), Hin recombinase (acting on hix sites), β recombinase (acting on six sites), Sin recombinase (acting on resH sites), Tn3 recombinase (acting on res

sites), γδ recombinase (acting on res sites), Cre recombinase from bacteriophage P1 (acting on LoxP sites); λ phage integrase (acting on att sites); FLP recombinases of fungal origin (acting on FTR sites); phiC31 integrase; Dre recombinase, BxB1; and prokaryotic β-recombinase.

[00116] For example, the evolved recombinase for use with the compositions and methods of the instant disclosure may have been altered to interact with (*e.g.*, bind and recombine) a non-canonical recombinase target sequence. As a non-limiting example, the non-canonical recombinase target sequence may be naturally occurring, such as, for example, sequences within a "safe harbor" genomic locus in a mammalian genome, *e.g.*, a genomic locus that is known to be tolerant to genetic modification without any undesired effects. Recombinases targeting such sequences allow, *e.g.*, for the targeted insertion of nucleic acid constructs at a specific genomic location without the need for conventional time- and labor-intensive gene targeting procedures, *e.g.*, via homologous recombination technology. In addition, the directed evolution strategies provided herein can be used to evolve recombinases with an altered activity profile, *e.g.*, recombinases that favor integration of a nucleic acid sequence over excision of that sequence or vice versa.

[00117] Evolved recombinases exhibit altered target sequence preferences as compared to their wild type counterparts, can be used to target virtually any target sequence for recombinase activity. Accordingly, the evolved recombinases can be used to modify, for example, any sequence within the genome of a cell or subject. Because recombinases can effect an insertion of a heterologous nucleic acid molecule into a target nucleic acid molecule, an excision of a nucleic acid sequence from a nucleic acid molecule, an inversion, or a replacement of nucleic acid sequences, the technology provided herein enables the efficient modification of genomic targets in a variety of ways (*e.g.*, integration, deletion, inversion, exchange of nucleic acid sequences).

[00118] Catalytic domains from evolved recombinases for use with the methods and compositions of the instant disclosure comprise an amino acid sequence that is at least 70%, at least 80%, at least 90%, at least 95%, or at least 97% identical to the sequence of a wild-type recombinase, wherein the amino acid sequence of the evolved recombinase comprises at least one mutation as compared to the sequence of the wild-type recombinase, and wherein the evolved recombinase recognizes a DNA recombinase target sequence that differs from the canonical recombinase target sequence by at least one nucleotide. In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence that differs from the canonical recombinase target sequence (*e.g.*, a res, gix, hix, six, resH, LoxP, FTR, or att core

or related core sequence) by at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20 at least 25, or at least 30 nucleotides. In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence that differs from the canonical recombinase target sequence by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides.

[00119] In some embodiments, only a portion of the recombinase is used in the fusion proteins and methods described herein. As a non-limiting embodiment, only the C-terminal portion of the recombinase may be used in the fusion proteins and methods described herein. In a specific embodiment, the 25 kDa carboxy-terminal domain of Cre recombinase may be used in the compositions and methods. See, for example, Hoess *et al*, "DNA Specificity of the Cre Recombinase Resides in the 25 kDa Carboxyl Domain of the Protein," J. Mol. Bio. 1990 Dec 20, 216(4):873-82, which is incorporated by reference herein for all purposes. The 25 kDa carboxy-terminal domain of Cre recombinase is the portion stretching from R118 to the carboxy terminus of the protein. In some embodiments, the 25kDa carboxy-terminal domain of Cre recombinase for use in the instant fusion proteins and methods may differ from the canonical 25kDa carboxy-terminal domain of Cre recombinase by at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, or at least 20 amino acids. In some embodiments, the 25kDa carboxy-terminal domain of Cre recombinase for use in the instant fusion proteins and methods may differ from the canonical 25kDa carboxy-terminal domain of Cre recombinase by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids. In certain embodiments, only a portion of the 25kDa carboxy-terminal domain of Cre recombinase may be used in the fusion proteins and methods described herein. For example, the portion of Cre recombinase used may be R130 to the carboxy terminus of the protein, T140 to the carboxy terminus of the protein, E150 to the carboxy terminus of the protein, N160 to the carboxy terminus of the protein, T170 to the carboxy terminus of the protein, I180 to the carboxy terminus of the protein, G190 to the carboxy terminus of the protein, T200 to the carboxy terminus of the protein, E210 to the carboxy terminus of the protein, L220 to the carboxy terminus of the protein, V230 to the carboxy terminus of the protein, C240 to the carboxy terminus of the protein, P250 to the carboxy terminus of the protein, A260 to the carboxy terminus of the protein, R270 to the carboxy terminus of the protein, G280 to the carboxy terminus of the protein, S290 to the carboxy terminus of the protein, A300 to the carboxy terminus of the protein, or M310 to the

carboxy terminus of the protein. As another set of non-limiting examples, the portion of Cre recombinase used may be R118-E340, R118-S330, R118-I320, R118-M310, R118-A300, R118-S290, R118-G280, R118-R270, R118-A260, R118-P250, R118-C240, R118-V230, R118-L220, or R118-E210. As a further set of non-limiting examples, the portion of Cre recombinase used may be R118-E210, G190-R270, E210-S290, P250-M310, or R270 to the carboxy terminus of the protein.

[00120] In some embodiments, the Cre recombinase used in the fusion proteins and methods described herein may be truncated at any position. In a specific embodiment, the Cre recombinase used in the fusion proteins and methods described herein may be truncated such that it begins with amino acid R118, A127, E138, or R154) (preceded in each case by methionine). In another set of non-limiting embodiments, the Cre recombinase used in the fusion proteins and methods described herein may be truncated within 10 amino acids, 9 amino acids, 8 amino acids, 7 amino acids, 6 amino acids, 5 amino acids, 4 amino acids, 3 amino acids, 2 amino acids, or 1 amino acid of R118, A127, E138, or R154.

[00121] In some embodiments, the recombinase target sequence is between 10-50 nucleotides long. In some embodiments, the recombinase is a Cre recombinase, a Hin recombinase, or a FLP recombinase. In some embodiments, the canonical recombinase target sequence is a LoxP site (5′-ATAACTTCGTATA GCATACAT TATACGAAGTTAT-3′ (SEQ ID NO: 739). In some embodiments, the canonical recombinase target sequence is an FRT site (5'-GAAGTTCCTATTCTCTAGAAA GTATAGGAACTTC -3′) (SEQ ID NO: 740). In some embodiments, the amino acid sequence of the evolved recombinase comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 mutations as compared to the sequence of the wild-type recombinase. In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence that comprises a left half-site, a spacer sequence, and a right half-site, and wherein the left half-site is not a palindrome of the right half-site.

[00122] In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence that comprises a naturally occurring sequence. In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence that is comprised in the genome of a mammal. In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence comprised in the genome of a human. In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence that occurs only once in the genome of a mammal. In some embodiments, the evolved

recombinase recognizes a DNA recombinase target sequence in the genome of a mammal that differs from any other site in the genome by at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 nucleotide(s). In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence located in a safe harbor genomic locus. In some embodiments, the safe harbor genomic locus is a Rosa26 locus. In some embodiments, the evolved recombinase recognizes a DNA recombinase target sequence located in a genomic locus associated with a disease or disorder.

[00123] In certain embodiments, the evolved recombinase may target a site in the Rosa locus of the human genome (*e.g.*, 36C6). A non-limiting set of such recombinases may be found, for example, in International PCT Publication, WO 2017/015545A1, published January 26, 2017, entitled "Evolution of Site Specific Recombinases," which is incorporated by reference herein for this purpose. In some embodiments, the amino acid sequence of the evolved recombinase comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 mutations as compared to the sequence of the wild-type recombinase. The nucleotide sequence encoding 36C6 is shown below in bold; those encoding GGS linkers are shown in italics; those encoding dCas9 linkers are black; those encoding the FLAG tag and NLS are underlined and in lowercase, respectively.

dCas9-36C6 (nucleotide) (SEQ ID NO: 765)

ATGTCCAACCTCCTTACCGTCCACCAGAATCTCCCTGCCCTTCCGGTGGATGCCACCTCTGATGAAGTGCGAAAA
AACCTGATGGATATGTTTCGCGATAGGCAAGCTTTTTCTGAACACACGTGGAAGATGCTCCTGTCAGTGTGTAGA
AGCTGGGCAGCTTGGTGCAAGTTGAACAACCGAAAATGGTTTCCTGCCGAACCCGAAGATGTGAGAGACTACCTC
CTCTACCTGCAGGCTCGAGGGCTCGCCGTGAAAACAATCCAACAACACTTGGGTCAGCTCAACATGCTGCACAGG
AGATCTGGGCTGCCCCGGCCGAGTGACTCTAATGCCGTTAGTCTCGTAATGCGGCGCATTCGCAAAGAGAATGTG
GATGCTGGAGAACGGGCGAAACAGGCACTGGCTTTTGAACGGACCGACTTCGATCAGGTGCGGAGTCTTATGGAG
AATAGTGACAGATGCCAGGACATTCGGAACCTTGCATTCCTGGGTATCGCGTATAATACCCTGCTGAGAATCGCT
GAGATCGCCAGAATCAGGGTAAAGGATATTTCTCGAACGGACGGGGGACGGATGTTGATTCATATCGGTCGCACT
AAAACACTTGTGAGTACCGCCGGGGTAGAGAAAGCCCTGAGCCTTGGAGTTACTAAACTGGTGGAGCGGTGGATT
AGCGTGTCCGGCGTGGCGGATGACCCAAACAATTACTTGTTTTGTAGGGTGCGGAAAAATGGTGTAGCCGCTCCA
TCCGCTACCTCACAGTTGAGTACACGCGCGTTGGAGGGGATTTTCGAAGCCACACATCGCTTGATCTACGGCGCC
AAGGACGATTCAGGCCAGCGATATCTTGCCTGGAGCGGGCATAGTGCCCGGGTGGGTGCCGCCCGAGACATGGCA
AGGGCTGGCGTGTCAATTCCTGAAATCATGCAGGCCGGCGGGTGGACCAACGTGAACATTGTGATGAACTATATC
CGGAACCTGGATAGCGAGACCGGAGCAATGGTCAGACTGCTTGAGGATGGCGAC*GGTGGATCCGGAGGGTCCGGA*
*GGTAGTGGCGGCAGCGGTGGTTCAGGTGGCAGCGGAGGGTCAGGAGGCTCT*GATAAAAAGTATTCTATTGGTTTA
GCTATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTG
TTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCA
GAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAA

ATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAG
GACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACG
ATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCC
CATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTG
TTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAG
GCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAA
AATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCT
GAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGAT
CAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAAT
ACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTT
CTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTAC
GCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGAT
GGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGC
ATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAA
GACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAAC
TCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAA
GGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCT
AAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATG
CGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTG
ACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAA
GATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAAC
GAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAA
AGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGG
GGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAG
AGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAA
AAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAA
AAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATT
GTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATA
GAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAG
AAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCT
GATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGC
TCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGG
CAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCT
GAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATA
CTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAG
TCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCG
CACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTG
TATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCC
AAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAA
CGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGA
AAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCG
ATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTC
GATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCA
GTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCG
AAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGC
CGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAAT
TTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTT
GAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGAT
GCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATT
ATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAA
CGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGG
ATAGATTTGTCACAGCTTGGGGGTGAC*GGTGGCTCC*GATTATAAGGATGATGACGACAAG*GGAGGTTCC*ccaaag
aagaaaaggaaggtcTGA

dCas9-36C6 (amino acid) (SEQ ID NO: 766)

MSNLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAEPEDVRDYL
LYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLME
NSDRCQDIRNLAFLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWI
SVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMA
RAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGD*GGSGGSGGSGGSGGSGGSGGSGGS*DKKYSIGL
AIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM
RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDN
EENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRK
RPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR
IDLSQLGGD*GGS*<u>DYKDDDDK</u>*GGS*pkkkrkv Stop

[00124] Some aspects of this disclosure provide evolved recombinases (*e.g.*, a Cre recombinase) comprising an amino acid sequence that is at least 70%, at least 80%, at least 90%, at least 95%, or at least 97% identical to the sequence of the recombinase sequence (*e.g.*, a Cre recombinase) discussed herein, wherein the amino acid sequence of the recombinase (*e.g.*, a Cre recombinase) comprises at least one mutation as compared to the sequence of the recombinase (*e.g.*, a Cre recombinase) sequence discussed herein, and wherein the recombinase (*e.g.*, a Cre recombinase) recognizes a DNA recombinase target sequence that differs from the canonical LoxP site 5′-ATAACTTCGTATA GCATACAT TATACGAAGTTAT-3′ (SEQ ID NO: 739) in at least one nucleotide.

[00125] In some embodiments, the amino acid sequence of the evolved recombinase (*e.g.*, a Cre recombinase) comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 mutations as compared to the sequence of the recombinase (*e.g.*, a Cre recombinase) sequence discussed herein and recognizes a DNA recombinase target sequence that differs

from the canonical target site (*e.g.*, a LoxP site) in at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 nucleotides.

[00126] In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence that comprises a left half-site, a spacer sequence, and a right half-site, wherein the left half-site is not a palindrome of the right half-site. In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence that comprises a naturally occurring sequence. In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence that is comprised in the genome of a mammal.

[00127] In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence that is comprised in the genome of a human. In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence that is comprised only once in the genome of a mammal. In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence in the genome of a mammal that differs from any other site in the genome by at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, or at least 15 nucleotide(s). In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence located in a safe harbor genomic locus. In some embodiments, the safe harbor genomic locus is a Rosa26 locus. In some embodiments, the evolved Cre recombinase recognizes a DNA recombinase target sequence located in a genomic locus associated with a disease or disorder.

[00128] Additional evolved recombinases (and methods for making the same) for use with the instant methods and compositions may be found in, for example, U.S. Patent Application No.: 15/216,844, which is incorporated herein by reference.

[00129] Additional suitable recombinases will be apparent to those of skill in the art for both providing recombinase catalytic domains or evolved recombinase catalytic domains, and such suitable recombinases include, without limitation, those disclosed in Hirano *et al.*, Site-specific recombinases as tools for heterologous gene integration. Appl Microbiol Biotechnol. 2011 Oct;92(2):227-39; Fogg *et al.*, New applications for phage integrases. J Mol Biol. 2014 Jul 29;426(15):2703; Brown *et al.*, Serine recombinases as tools for genome engineering. Methods. 2011 Apr;53(4):372-9; Smith *et al.*, Site-specific recombination by phiC31 integrase and other large serine recombinases. Biochem Soc Trans. 2010 Apr;38(2):388-94; Grindley *et al.*, Mechanisms of site-specific recombination. Annu Rev

Biochem. 2006;75:567-605; Smith *et al.*, Diversity in the serine recombinases. Mol Microbiol. 2002 Apr;44(2):299-307; Grainge *et al.*, The integrase family of recombinase: organization and function of the active site. Mol Microbiol. 1999 Aug;33(3):449-56; Gopaul *et al.*, Structure and mechanism in site-specific recombination. Curr Opin Struct Biol. 1999 Feb;9(1):14-20; Cox *et al.*, Conditional gene expression in the mouse inner ear using Cre-loxP. J Assoc Res Otolaryngol. 2012 Jun;13(3):295-322; Birling *et al.*, Site-specific recombinases for manipulation of the mouse genome. Methods Mol Biol. 2009;561:245-63; and Mishina M, Sakimura K. Conditional gene targeting on the pure C57BL/6 genetic background. Neurosci Res. 2007 Jun;58(2):105-12; the entire contents of each of which are incorporated herein by reference.

Structure of the Fusion Protein

[00130] The fusion protein of the instant instant disclosure may be any combination and order of the elements described herein. Exemplary fusion proteins include, but are not limited to, any of the following structures: NH_2-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein has the structure NH_2-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein has the structure NH_2-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[affinity tag]-COOH. In another embodiment, the fusion protein has the structure NH_2-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[NLS domain]-[linker sequence]-[affinity tag]-COOH.

[00131] In another embodiment, the fusion protein has the structure NH_2-[recombinase catalytic domain]-[optional linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[affinity tag]-COOH, NH_2-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH, NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-

programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]- [optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH, NH_2-[affinity tag]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH, NH_2-[affinity tag]-[optional linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-COOH, or NH_2-[affinity tag]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH.

[00132] In another embodiment, the fusion protein has the structure: NH_2-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[affinity tag]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[NLS domain]-[linker sequence]-[affinity tag]-COOH.

[00133] In another embodiment, the fusion protein has the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another

embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[NLS domain]-[linker sequence]-[affinity tag]-COOH.

[00134] In another embodiment, the fusion protein has the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker

sequence]-[affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[NLS domain]-[linker sequence]-[affinity tag]-COOH.

[00135] In another embodiment, the fusion protein has the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[NLS domain]-[linker sequence]-[affinity tag]-COOH.

[00136] In another embodiment, the fusion protein has the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-

programmable DNA binding protein domain]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[affinity tag]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[NLS domain]-[linker sequence]-[affinity tag]-COOH.

[00137] In one embodiment, the fusion protein has the structure NH_2-[optional affinity tag]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[optional affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[linker sequence]-[NLS domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-COOH.

[00138] In one embodiment, the fusion protein has the structure NH_2-[optional affinity tag]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[guide

nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[optional affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-COOH. In one embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[linker sequence]-[NLS domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-COOH.

[00139] In another embodiment, the fusion protein has the structure NH_2-[optional affinity tag]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[optional affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[linker sequence]-[NLS domain]-[linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH.

[00140] In another embodiment, the fusion protein has the structure NH_2-[optional affinity tag]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[optional affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[linker sequence]-[NLS domain]-[linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH.

[00141] In another embodiment, the fusion protein has the structure NH_2-[optional affinity tag]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[optional affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide

sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[linker sequence]-[NLS domain]-[linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH.

[00142] In another embodiment, the fusion protein has the structure NH_2-[optional affinity tag]-[optional linker sequence]-[optional NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[optional affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH. In another embodiment, the fusion protein comprises the structure NH_2-[affinity tag]-[linker sequence]-[NLS domain]-[linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[linker sequence]-[C-terminal portion of a

bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-COOH.

[00143] The fusion protein may further comprise one or more affinity tags. Suitable affinity tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, polyarginine (poly-Arg) tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. The FLAG tag may have the sequence PKKKRKV (SEQ ID NO: 702). The one or more affinity tags are bound to the guide nucleotide sequence-programmable DNA binding protein domain, the recombinase catalytic domain, or the NLS domain via one or more third linkers. The third linker may be any peptide linker described herein. For example, the third linker may be a peptide linker.

[00144] As a non-limiting set of examples, the third linker may comprise an XTEN linker SGSETPGTSESATPES (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9), an amino acid sequence comprising one or more repeats of the tri-peptide GGS, or any of the following amino acid sequences: VPFLLEPDNINGKTC (SEQ ID NO: 10), GSAGSAAGSGEF (SEQ ID NO: 11), SIVAQLSRPDPA (SEQ ID NO: 12), MKIIEQLPSA (SEQ ID NO: 13), VRHKLKRVGS (SEQ ID NO: 14), GHGTGSTGSGSS (SEQ ID NO: 15), MSRPDPA (SEQ ID NO; 16), or GGSM (SEQ ID NO: 17). In certain embodiments, the third linker comprises one or more repeats of the tri-peptide GGS. In an embodiment, the third linker comprises from one to five repeats of the tri-peptide GGS. In another embodiment, the third linker comprises one repeat of the tri-peptide GGS. In a specific embodiment, the third linker has the sequence GGS.

[00145] The third linker may also be a non-peptide linker. In certain embodiments, the non-peptide linker comprises polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacryl amide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker. In other embodiments, the alkyl linker has the formula: —NH—$(CH_2)_s$—C(O)—, wherein s may be any integer between 1 and 100, inclusive. In a specific embodiment, s is any integer between 1 and 20, inclusive.

[00146] The fusion protein of the instant disclosure has greater than 90%, 95%, or 99% sequence identity with the amino acid sequence shown in amino acids 1-1544 of SEQ ID NO: 185, which is identical to the sequence shown in SEQ ID NO: 719.

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQ KGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYV MGALAEMERELIIERTMAGLAAARNKGRRFGRPPK*GGSGGSGGSGGSGGSGGSG GSGGS***DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL IEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD***GGS***<u>DYK DDDDK</u> Stop (SEQ ID NO: 719)

[00147] In the context of proteins that dimerize (or multimerize) such as, for example, fusions between a nuclease-inactivated Cas9 (or a Cas9 gRNA binding domain) and a recombinase (or catalytic domain of a recombinase), a target site typically comprises a left-half site (bound by one protein), a right-half site (bound by the second protein), and a spacer sequence between the half sites in which the recombination is made. In some embodiments, either the left-half site or the right half-site (and not the spacer sequence) is recombined. In other embodiments, the spacer sequence is recombined. This structure ([left-half site]-[spacer sequence]-[right-half site]) is referred to herein as an LSR structure. In some embodiments, the left-half site and/or the right-half site correspond to an RNA-guided target site (*e.g.,* a Cas9 target site). In some embodiments, either or both half-sites are shorter or longer than *e.g.,* a typical region targeted by Cas9, for example shorter or longer than 20 nucleotides. In some embodiments, the left and right half sites comprise different nucleic acid sequences. In some embodiments, the spacer sequence is at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 125, at least 150, at least 175, at least 200, or at least 250 bp long. In some embodiments, the spacer sequence is between approximately 15 bp and approximately 25 bp long. In some embodiments, the spacer sequence is approximately 15 bp long. In some embodiments, the spacer sequence is approximately 25 bp long.

EXAMPLES

Example 1: A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells

Materials and Methods

Oligonucleotides and PCR

[00148] All oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, CA) and are listed in Tables 1-5. Enzymes, unless otherwise noted, were purchased from New England Biolabs (Ipswich, MA). Plasmid Safe ATP-dependent DNAse was purchased from Epicentre (Madison, WI). All assembled vectors were transformed into One Shot Mach1-T1 phage-resistant chemically competent cells (Fisher Scientific, Waltham, MA). Unless otherwise noted, all PCR reactions were performed with Q5 Hot Start High-Fidelity 2X Master Mix. Phusion polymerse was used for circular polymerase extension cloning (CPEC) assemblies.

Table 1. Oligonucleotides for gRNA construction

Oligonucleotide Name	Sequence	SEQ ID NO:
R.pHU6.TSS(-1).univ	GGTGTTTCGTCCTTTCCACAAG	20
F.non-target	GCACACTAGTTAGGGATAACAGTTTTAG AGCTAGAAATAGC	21
F.Chr10-1	GCCCATGACCCTTCTCCTCTGTTTTAGAG CTAGAAATAGC	22
F.Chr10-1-rev	GCTCAGGGCCTGTGATGGGAGGTTTTAG AGCTAGAAATAGC	23
F.Chr10-2	GGCCCATGACCCTTCTCCTCGTTTTAGAG CTAGAAATAGC	24
F.Chr10-2rev	GCCTCAGGGCCTGTGATGGGAGTTTTAG AGCTAGAAATAGC	25
F.Centromere_Chr_1_5_19-gRNA-for	GACTTGAAACACTCTTTTTCGTTTTAGAG CTAGAAATAGC	26
F.Centromere_Chr_1_5_19-gRNA-rev	GAGTTGAAGACACACAACACAGTTTTAG AGCTAGAAATAGC	27
F.Ch5_155183064-gRNA-for	GGAACTCATGTGATTAACTGGTTTTAGA GCTAGAAATAGC	28
F.Ch5_155183064-gRNA-rev-1	GTCTACCTCTCATGAGCCGGTGTTTTAGA GCTAGAAATAGC	29
F.Ch5_169395198-gRNA-for	GTTTCCCGCAGGATGTGGGATGTTTTAG AGCTAGAAATAGC	30
F.Ch5_169395198-gRNA-rev	GCCTGGGGATTTATGTTCTTAGTTTTAGA GCTAGAAATAGC	31
F.Ch12_62418577-gRNA-for	GAAATAGCACAATGAATGGAAGTTTTAG AGCTAGAAATAGC	32
F.Ch12_62418577-gRNA-rev	GACTTTTTGGGGGAGAGGGAGGTTTTAG AGCTAGAAATAGC	33
F.Ch13_102010574-gRNA-for	GGAGACTTAAGTCCAAAACCGTTTTAGA GCTAGAAATAGC	34
F.Ch13_102010574-gRNA-rev	GTCAGCTATGATCACTTCCCTGTTTTAGA GCTAGAAATAGC	35

Table 2. Oligonucleotides and gBlocks for reporter construction

Construct Name	Sequence	SEQ ID NO:
1-0bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTCTGTAAACCGAGGTGAGACGG	36
1-0bp-rev	CCGTCTCACCTCGGTTTACAGACCTCTGTTTGG GAAAATTGGGGACGCCGAGACGA	37
1-1bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTtCTGTAAACCGAGGTGAGACGG	38
1-1bp-rev	CCGTCTCACCTCGGTTTACAGaACCTCTGTTTGG GAAAATTGGGGACGCCGAGACGA	39
1-2bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTatCTGTAAACCGAGGTGAGACGG	40
1-2bp-rev	CCGTCTCACCTCGGTTTACAGatACCTCTGTTTG GGAAAATTGGGGACGCCGAGACGA	41
1-3bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTaatCTGTAAACCGAGGTGAGACGG	42
1-3bp-rev	CCGTCTCACCTCGGTTTACAGattACCTCTGTTTG GGAAAATTGGGGACGCCGAGACGA	43
1-4bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTaaatCTGTAAACCGAGGTGAGACGG	44
1-4bp-rev	CCGTCTCACCTCGGTTTACAGatttACCTCTGTTT GGGAAAATTGGGGACGCCGAGACGA	45
1-5bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTgaaatCTGTAAACCGAGGTGAGACGG	46
1-5bp-rev	CCGTCTCACCTCGGTTTACAGatttcACCTCTGTTT GGGAAAATTGGGGACGCCGAGACGA	47
1-6bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTcgaaatCTGTAAACCGAGGTGAGACGG	48
1-6bp-rev	CCGTCTCACCTCGGTTTACAGatttcgACCTCTGTT TGGGAAAATTGGGGACGCCGAGACGA	49
1-7bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTtcgaaatCTGTAAACCGAGGTGAGACGG	50
1-7bp-rev	CCGTCTCACCTCGGTTTACAGatttcgaACCTCTGT TTGGGAAAATTGGGGACGCCGAGACGA	51
2-0bp-for	TCGTCTCGGAGGTTTTGGAACCTCTGTTTGGGA AAATTGGGGAGTCTGAGACGG	52
2-0bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTTCCAAAACCTCCGAGACGA	53
2-1bp-for	TCGTCTCGGAGGTTTTGGACACCTCTGTTTGGG AAAATTGGGGAGTCTGAGACGG	54
2-1bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTGTCCAAAACCTCCGAGACGA	55
2-2bp-for	TCGTCTCGGAGGTTTTGGACTACCTCTGTTTGG GAAAATTGGGGAGTCTGAGACGG	56
2-2bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTAGTCCAAAACCTCCGAGACGA	57

2-3bp-for	TCGTCTCGGAGGTTTTGGACTTACCTCTGTTTG GGAAAATTGGGGAGTCTGAGACGG	58
2-3bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTAAGTCCAAAACCTCCGAGACGA	59
2-4bp-for	TCGTCTCGGAGGTTTTGGACTTAACCTCTGTTT GGGAAAATTGGGGAGTCTGAGACGG	60
2-4bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTTAAGTCCAAAACCTCCGAGACGA	61
2-5bp-for	TCGTCTCGGAGGTTTTGGACTTAGACCTCTGTT TGGGAAAATTGGGGAGTCTGAGACGG	62
2-5bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTCTAAGTCCAAAACCTCCGAGACGA	63
2-6bp-for	TCGTCTCGGAGGTTTTGGACTTAGCACCTCTGT TTGGGAAAATTGGGGAGTCTGAGACGG	64
2-6bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTGCTAAGTCCAAAACCTCCGAGACGA	65
2-7bp-for	TCGTCTCGGAGGTTTTGGACTTAGCTACCTCTG TTTGGGAAAATTGGGGAGTCTGAGACGG	66
2-7bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTAGCTAAGTCCAAAACCTCCGAGACGA	67
4-0bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTCTGTAAACCGATGAGACGG	68
4-0bp-rev	CCGTCTCATCGGTTTACAGACCTCTGTTTGGGA AAATTGGGGGTGCAGAGACGA	69
4-1bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTtCTGTAAACCGATGAGACGG	70
4-1bp-rev	CCGTCTCATCGGTTTACAGaACCTCTGTTTGGG AAAATTGGGGGTGCAGAGACGA	71
4-2bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTatCTGTAAACCGATGAGACGG	72
4-2bp-rev	CCGTCTCATCGGTTTACAGatACCTCTGTTTGGG AAAATTGGGGGTGCAGAGACGA	73
4-3bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTaatCTGTAAACCGATGAGACGG	74
4-3bp-rev	CCGTCTCATCGGTTTACAGattACCTCTGTTTGGG AAAATTGGGGGTGCAGAGACGA	75
4-4bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTaaatCTGTAAACCGATGAGACGG	76
4-4bp-rev	CCGTCTCATCGGTTTACAGatttACCTCTGTTTGG GAAAATTGGGGGTGCAGAGACGA	77
4-5bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTgaaatCTGTAAACCGATGAGACGG	78
4-5bp-rev	CCGTCTCATCGGTTTACAGatttcACCTCTGTTTGG GAAAATTGGGGGTGCAGAGACGA	79
4-6bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTcgaaatCTGTAAACCGATGAGACGG	80
4-6bp-rev	CCGTCTCATCGGTTTACAGatttcgACCTCTGTTTG GGAAAATTGGGGGTGCAGAGACGA	81

4-7bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTtcgaaatCTGTAAACCGATGAGACGG	82
4-7bp-rev	CCGTCTCATCGGTTTACAGatttcgaACCTCTGTTT GGGAAAATTGGGGGTGCAGAGACGA	83
5-0bp-for	TCGTCTCGCCGAGGTTTTGGAACCTCTGTTTGG GAAAATTGGGGCTCGTGAGACGG	84
5-0bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTTCCAAAACCTCGGCGAGACGA	85
5-1bp-for	TCGTCTCGCCGAGGTTTTGGACACCTCTGTTTG GGAAAATTGGGGCTCGTGAGACGG	86
5-1bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTGTCCAAAACCTCGGCGAGACGA	87
5-2bp-for	TCGTCTCGCCGAGGTTTTGGACTACCTCTGTTT GGGAAAATTGGGGCTCGTGAGACGG	88
5-2bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTAGTCCAAAACCTCGGCGAGACGA	89
5-3bp-for	TCGTCTCGCCGAGGTTTTGGACTTACCTCTGTT TGGGAAAATTGGGGCTCGTGAGACGG	90
5-3bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTAAGTCCAAAACCTCGGCGAGACGA	91
5-4bp-for	TCGTCTCGCCGAGGTTTTGGACTTAACCTCTGT TTGGGAAAATTGGGGCTCGTGAGACGG	92
5-4bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTTAAGTCCAAAACCTCGGCGAGACGA	93
5-5bp-for	TCGTCTCGCCGAGGTTTTGGACTTAGACCTCTG TTTGGGAAAATTGGGGCTCGTGAGACGG	94
5-5bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTCTAAGTCCAAAACCTCGGCGAGACGA	95
5-6bp-for	TCGTCTCGCCGAGGTTTTGGACTTAGCACCTCT GTTTGGGAAAATTGGGGCTCGTGAGACGG	96
5-6bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTGCTAAGTCCAAAACCTCGGCGAGACGA	97
5-7bp-for	TCGTCTCGCCGAGGTTTTGGACTTAGCTACCTC TGTTTGGGAAAATTGGGGCTCGTGAGACGG	98
5-7bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGAG GTAGCTAAGTCCAAAACCTCGGCGAGACGA	99
1-Chr10--54913298-54913376-for	TCGTCTCGGCGTCCCCTCCCATCACAGGCCCTG AGGTTTAAGAGAAAACCTGAGACGG	100
1-Chr10-54913298-54913376-rev	CCGTCTCAGGTTTTCTCTTAAACCTCAGGGCCT GTGATGGGAGGGGACGCCGAGACGA	101
2-Chr10--54913298-54913376-for	TCGTCTCGAACCATGGTTTTGTGGGCCAGGCCC ATGACCCTTCTCCTCTGGGAGTCTGAGACGG	102
2-Chr10--54913298-54913376-rev	CCGTCTCAGACTCCCAGAGGAGAAGGGTCATG GGCCTGGCCCACAAAACCATGGTTCGAGACGA	103
4-Chr10-54913298-54913376-for	TCGTCTCTGCACCCCCTCCCATCACAGGCCCTG AGGTTTAAGAGAAAACCATTGAGACGG	104
4-Chr10-54913298-54913376-rev	CCGTCTCAATGGTTTTCTCTTAAACCTCAGGGC CTGTGATGGGAGGGGGTGCAGAGACGA	105
5-Chr10-54913298-	TCGTCTCGCCATGGTTTTGTGGGCCAGGCCCAT	106

54913376-for	GACCCTTCTCCTCTGGGCTCGTGAGACGG	
5-Chr10-54913298-54913376-rev	CCGTCTCACGAGCCCAGAGGAGAAGGGTCATGGGCCTGGCCCACAAAACCATGGCGAGACGA	107
3-for	ATCCGTCTCCAGTCGAGTCGGATTTGATCTGATCAAGAGACAG	108
3-rev	AACCGTCTCGGTGCGTTCGGATTTGATCCAGACATGATAAGATAC	109
Esp3I-insert-for	/Phos/CGCGTTGAGACGCTGCCATCCGTCTCGC	110
Esp3I-insert-rev	/Phos/TCGAGCGAGACGGATGGCAGCGTCTCAA	111
Centromere_Chr_1_5_19-1_2*	GTTGTTCGTCTCGGCGTCCTTGTGTTGTGTGTCTTCAACTCACAGAGTTAAACGATGCTTTACACAGAGTAGACTTGAAACACTCTTTTTCTGGAGTCTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	112
Centromere_Chr_1_5_19-4_5*	GTTGGTCGTCTCTGCACCCTTGTGTTGTGTGTCTTCAACTCACAGAGTTAAACGATGCTTTACACAGAGTAGACTTGAAACACTCTTTTTCTGGCTCGTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	113
Ch5_155183064-155183141-1_2*	GTTGTTCGTCTCGGCGTCCCACCGGCTCATGAGAGGTAGAGCTAAGGTCCAAACCTAGGTTTATCTGAGACCGGAACTCATGTGATTAACTGTGGAGTCTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	114
Ch5_155183064-155183141-4_5*	GTTGGTCGTCTCTGCACCCCACCGGCTCATGAGAGGTAGAGCTAAGGTCCAAACCTAGGTTTATCTGAGACCGGAACTCATGTGATTAACTGTGGCTCGTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	115
Ch5_169395198-169395274-1_2*	GTTGTTCGTCTCGGCGTCCTTAAGAACATAAATCCCCAGGAATTCACAGAAACCTTGGTTTGAGCTTTGGATTTCCCGCAGGATGTGGGATAGGAGTCTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	116
Ch5_169395198-169395274-4_5*	GTTGGTCGTCTCTGCACCCTTAAGAACATAAATCCCCAGGAATTCACAGAAACCTTGGTTTGAGCTTTGGATTTCCCGCAGGATGTGGGATAGGCTCGTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	117
Ch12_62418577-62418652-1_2*	GTTGTTCGTCTCGGCGTCCACTCCCTCTCCCCCAAAAAGTAAAGGTAGAAAACCAAGGTTTACAGGCAACAAATAGCACAATGAATGGAATGGAGTCTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	118
Ch12_62418577-62418652-4_5*	GTTGGTCGTCTCTGCACCCACTCCCTCTCCCCCAAAAAGTAAAGGTAGAAAACCAAGGTTTACAGGCAACAAATAGCACAATGAATGGAATGGCTCGTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	119
chr13_102010574-102010650-1_2*	GTTGTTCGTCTCGGCGTCCTAGGGAAGTGATCATAGCTGAGTTTCTGGAAAAACCTAGGTTTTAAAGTTGAGGAGACTTAAGTCCAAAACCTGGAGTCTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	120
chr13_102010574-102010650-4_5*	GTTGGTCGTCTCTGCACCCTAGGGAAGTGATCATAGCTGAGTTTCTGGAAAAACCTAGGTTTTAAAGTTGAGGAGACTTAAGTCCAAAACCTGGCTCG	121

	TGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	
Oligonucleotide sequences were annealed to create the fragments shown in Figure 1. The names correspond to the fragment number (1, 2, 4, or 5) and then to the number of base pair spacer nucleotides separating the Cas9 binding site from the gix core site. * Double stranded gBlocks as described in the methods within the supporting material document.		

Table 3. Oligonucleotides for recCas9 construction

Oligonucleotide Name	Sequence	SEQ ID NO:
1GGS-link-for_BamHI	TTCATCGGATCCGATAAAAAGTATTCTATTGGTTTAGCTATCGGCAC	122
5GGS-link-for_BamHI	TTCATCGGATCCGGTGGTTCAGGTGGCAGCGGAG	123
8GGS-link-for_BamHI	TTCATCGGATCCGGAGGGTCCGGAGGTAGTGGCGGCAGCGGTGGTTCAGGTGGCAGCGGAG	124
Cas9-rev-FLAG-NLS-AgeI	AATAACCGGTTCAGACCTTCCTTTTCTTCTTTGGGGAACCTCCCTTGTCGTCATCATCCTTATAATCGGAGCCACCGTCACCCCCAAGCTGTGACAAATC	125
1GGS-rev-BamHI	TGATAAGGATCCACCCTTTGGTGGTCTTCCAAACCGCC	126
2GGS-rev-BamH	TGATAAGGATCCACCGCTACCACCCTTTGGTGGTCTTC	127
Gin-for_NotI	AGATCCGCGGCCGCTAATAC	128
Esp3I-for-plasmid	TTGAGTcgtctcTATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG	129
Esp3I-rev-plasmid	CTGGAAcgtctcACTGTCAGACCAAGTTTACTCATATATACTTTAGATTG	130
spec-Esp3I-for	GGTGTGcgtctcTACAGTTATTTGCCGACTACCTTGGTGATCTCGC	131
spec-Esp3I-rev	ACACCAcgtctcTGTATGAGGGAAGCGGTGATCGCC	132
cpec assembly-for-plasmid	CATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG	133
cpec assembly-rev-plasmid	CTGTCAGACCAAGTTTACTCATATATACTTTAGATTG	134
cpec assembly-for-spec	CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTTGCCGACTACCTTGGTGATCTCG	135
cpec assembly-for-spec2	CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTTGCCGACTACCTTGGTGATCTCG	136
cpec assembly-rev-spec	CCCTGATAAATGCTTCAATAATATTGAAAA	137

	AGGAAGAGTATG	

Table 4. Custom sequencing oligonucleotides

Oligonucleotide Name	Sequence	SEQ ID NO:
Fwd CMV	CGCAAATGGGCGGTAGGCGTG	138
Cas9coRevE1	CCGTGATGGATTGGTGAATC	139
Cas9coRevE2	CCCATACGATTTCACCTGTC	140
Cas9coRevE3	GGGTATTTTCCACAGGATGC	141
Cas9coRevE4	CTTAGAAAGGCGGGTTTACG	142
Cas9coRevE5	CTTACTAAGCTGCAATTTGG	143
Cas9coRevE6	TGTATTCATCGGTTATGACAG	144
bGH_PArev seq1	CAGGGTCAAGGAAGGCACG	145
pHU6-gRNA_for	GTTCCGCGCACATTTCC	146
pHU6-gRNA_rev	GCGGAGCCTATGGAAAAAC	147
pCALNL-for1	GCCTTCTTCTTTTTCCTACAGC	148
pCALNL-for2	CGCATCGAGCGAGCAC	149

Table 5. Genomic PCR primers

Oligonucleotide Name	Sequence	SEQ ID NO:
FAM19A2-F1	TCAAGTAGCAAAAGAAGTAGGAGTCAG	150
FAM19A2-F2	TTAGATGCATTCGTGCTTGAAG	151
FAM19A2-C1	TTAATTTCTGCTGCTAGAACTAAATCTGG	152
FAM19A2-R1	GGGAAGAAAACTGGATGGAGAATG	153
FAM19A2-R2	CATAAATGACCTAGTGGAGCTG	154
FAM19A2-C2	TGGTTATTTTGCCCATTAGTTGATGC	155

Reporter Construction

[00149] A five-piece Golden Gate assembly was used to construct reporters described below. Fragments 1-5 were flanked by Esp3I sites; Esp3I digestion created complementary 5′ overhangs specifying the order of fragment assembly (Figure 6). Fragments 1, 2, 4, and 5 were created by annealing forward and reverse complementary oligonucleotides listed in Table 5. Fragments were annealed by mixing 10 µl of each oligonucleotide (100 µM) in 20 µl of molecular grade water, incubating at 95 °C for 3 minutes and reducing the temperature to 16 °C at a rate of -0.1 °C/sec. Fragment 3 was created by PCR amplifying the region containing kanR and a PolyA stop codon with primers 3-for and 3-rev. These primers also appended Esp3I on the 5' and 3' ends of this sequence.

[00150] Annealed fragments 1, 2, 4 and 5 were diluted 12,000 fold and 0.625 µl of each fragment were added to a mixture containing the following:

1) 40-50 ng fragment 3
2) 100 ng pCALNL EGFP-Esp3I
3) 1 µL Tango Buffer (10X)
4) 1 µL DTT (10 mM)
5) 1 µL ATP (10 mM)
6) 0.25 uL T7 ligase (3,000 U/µL)
7) 0.75 uL Esp3I (10 U/µL)
8) H_2O up to 10 µL

[00151] Reactions were incubated in thermal cycler programmed for 20 cycles (37 °C for 5 min, 20 °C).

[00152] After completion of the Golden Gate reactions, 7 µL of each reaction was mixed with 1 µL of ATP (10 mM), 1 µL of 10X Plasmid Safe ATP-dependent DNAse buffer (10X), and 1 µL of Plasmid Safe ATP-dependent DNAse (10 U/µL) (Epicentre, Madison, WI) to remove linear DNA and reduce background. DNAse digestions were incubated at 37 °C for 30 min and heat killed at 70 °C for 30 min. Half (5 µL) of each reaction was transformed into Mach1-T1 cells. Colonies were analyzed by colony PCR and sequenced.

[00153] The protocol was modified for reporters used in Figure 4. Two gBlocks, encoding target sites to the 5′ or 3′ of the PolyA terminator were used instead of fragments 1, 2, 4 and 5. These gBlocks (10 ng) were added to the MMX, which was cycled 10 times (37 °C for 5 min, 20 °C) and carried forward as described above.

Plasmids

[00154] Unless otherwise stated, DNA fragments were isolated from agarose gels using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and further purified using DNA Clean & Concentrator-5 (Zymo Research, Irvine, CA) or Qiaquick PCR purification kit (Qiagen, Valencia, CA). PCR fragments not requiring gel purification were isolated using one of the kits listed above.

[00155] The pCALNL-GFP subcloning vector, pCALNL-EGFP-Esp3I, was used to clone all recCas9 reporter plasmids and was based on the previously described pCALNL-GFP vector (Matsuda and Cepko, Controlled expression of transgenes introduced by in vivo electroporation. Proceedings of the National Academy of Sciences of the United States of America 104, 1027-1032 (2007), which is incorporated herein by reference). To create pCALNL-EGFP-Esp3I, pCALNL-GFP vectors were digested with XhoI and MluI and gel

purified to remove the loxP sites, the kanamycin resistance marker, and the poly-A terminator. Annealed oligonucleotides formed an EspI-Insert, that contained inverted Esp3I sites as well as XhoI and MluI compatible overhangs; this insert was ligated into the XhoI and MluI digested plasmid and transformed.

[00156] pCALNL-GFP recCas9 reporter plasmids were created by Golden Gate assembly with annealed oligos and PCR products containing compatible Esp3I overhangs. Golden Gate reactions were set up and performed as described previously with Esp3I (ThermoFisher Scientific, Waltham, MA) (Sanjana *et al.*, A transcription activator-like effector toolbox for genome engineering. Nature protocols 7, 171-192 (2012), the entire contents of which is hereby incorporated by reference). Figure 6 outlines the general assembly scheme and relevant primers for reporter assembly as well as sequences for all recCas9 target sites are listed in Tables 2 and 6, respectively. A representative DNA sequence containing KanR (bold and underlined) and PolyA terminator (in italics and underlined) flanked by two recCas9 target sites is shown below. The target sites shown are both PAM_NT1-0bp-gix_core-0bp-NT1_PAM (see Table 6). Protoadjacent spacer motifs (PAMs) are in bold. Base pair spacers are lower case. Gix site or gix-related sites are in italics and dCas9 binding sites are underlined. For the genomic reporter plasmids used in the assays of Figure 4, a G to T transversion was observed in the kanamycin resistance marker, denoted by a G/T in the sequence below. This was present in all the reporters used in this figure, and it is not expected to affect the results, as it is far removed from the PolyA terminator and recCas9 target sites.

ACGCGT**CCC**CAATTTTCCCAAACAGAGGT*CTGTAAACCGAGGTTTTGGA*ACCTCTG
TTTGGGAAAATTG**GGG**AGTCGAGTCGGATTTGATCTGATCAAGAGACAGGATGA
GGATCGTTTCGC**ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCG**
CTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCT
GCTCTGATGCCGCCGTGTTCCGGCTGTCAG/TCGCAGGGGCGCCCGGTTCTT
TTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCA
GCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTC
GACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCG
GGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCA
TGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCAT
TCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAG
CCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC

CAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATC TCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATG GCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCT ATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCG AATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCA GCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGG GTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCC ACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCT GGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCATCGA TAAC*TTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCAC AAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCT TATC*ATGTCTGGATCAAATCCGAACGCAC**CCC**CAATTTTCCCAAACAGAGGT*CTG TAAACCGAGGTTTTGGA*ACCTCTGTTTGGGAAAATTG**GGG**CTCGAG (SEQ ID NO: 156)

Table 6. List of target site sequences used in reporter assays

Target site name	Sequence	SEQ ID NO:
PAM_NT1-0bp-gix_core-0bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTt*CTGTAAACCGAG GTTTTGG*AACCTCTGTTTGGGAAAATTG**GGG**	157
PAM_NT1-1bp-gix_core-1bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTt*CTGTAAACCGAG GTTTTGG*cAACCTCTGTTTGGGAAAATTG**GGG**	158
PAM_NT1-2bp-gix_core-2bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTat*CTGTAAACCGA GGTTTTGG*ctAACCTCTGTTTGGGAAAATTG**GGG**	159
PAM_NT1-3bp-gix_core-3bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTaat*CTGTAAACCG AGGTTTTGG*cttAACCTCTGTTTGGGAAAATTG**GGG**	160
PAM_NT1-4bp-gix_core-4bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTaaat*CTGTAAACCG AGGTTTTGG*cttaAACCTCTGTTTGGGAAAATTG**GGG**	161
PAM_NT1-5bp-gix_core-5bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTgaaat*CTGTAAACC GAGGTTTTGG*cttagAACCTCTGTTTGGGAAAATTG**G GG**	162
PAM_NT1-6bp-gix_core-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC CGAGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG **GGG**	163
PAM_NT1-7bp-gix_core-7bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTtcgaaat*CTGTAAAC CGAGGTTTTGG*cttagctAACCTCTGTTTGGGAAAATT G**GGG**	164

PAM_NT1-6bp-gix_core-0bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTtcgaaat*CTGTAAAC CGAGGTTTTGG*AACCTCTGTTTGGGAAAATTG**GGG**	165
PAM_NT1-6bp-gix_core-1bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTtcgaaat*CTGTAAAC CGAGGTTTTGG*cAACCTCTGTTTGGGAAAATTG**GG G**	166
PAM_NT1-6bp-gix_core-2bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC CGAGGTTTTGG*ctAACCTCTGTTTGGGAAAATTG**GG G**	167
PAM_NT1-6bp-gix_core-4bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC CGAGGTTTTGG*cttaAACCTCTGTTTGGGAAAATTG**G GG**	168
PAM_NT1-6bp-gix_core-5bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC CGAGGTTTTGG*cttagAACCTCTGTTTGGGAAAATTG **GGG**	169
PAM_NT1-0bp-gix_core-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGT*CTGTAAACCGAG GTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**GGG**	170
PAM_NT1-1bp-gix_core-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTt*CTGTAAACCGAG GTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**GGG**	171
PAM_NT1-2bp-gix_core-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTat*CTGTAAACCGA GGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**GG G**	172
PAM_NT1-3bp-gix_core-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTaat*CTGTAAACCG AGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**G GG**	173
PAM_NT1-4bp-gix_core-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTaaat*CTGTAAACCG AGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**G GG**	174
PAM_NT1-5bp-gix_core-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTgaaat*CTGTAAACC GAGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**G GG**	175
Chromosome_10-54913298-54913376*	**CCC**CTCCCATCACAGGCCCTGAGgtttaa*GAGAAAAC CATGGTTTTGTG*ggccagGCCCATGACCCTTCTCCTCT **GGG**	176
Centromere_Chromosomes_1_5_19	**CCT**TGTGTTGTGTGTCTTCAACTcacag*AGTTAAACGA TGCTTTACAC*agagtaGACTTGAAACACTCTTTTTC**TG G**	177
Chromosome_5_155183064-155183141 (site 1)	**CCA**CCGGCTCATGAGAGGTAGAGctaag*GTCCAAAC CTAGGTTTATCT*gagaccGGAACTCATGTGATTAACTG **TGG**	178
Chromosome_5_169395198-169395274 (site 2)	**CCT**TAAGAACATAAATCCCCAGGaattc*ACAGAAACC TTGGTTTGAGC*tttggaTTTCCCGCAGGATGTGGGAT**A GG**	179
Chromosome_12_62418577-62418652	**CCA**CTCCCTCTCCCCCAAAAAGTaaagg*TAGAAAACC AAGGTTTACAG*gcaacAAATAGCACAATGAATGGAA **TGG**	180
Chromosome_13_102	**CCT**AGGGAAGTGATCATAGCTGAgtttct*GGAAAAAC*	181

010574-102010650 (FGF14)	*CTAGGTTTTAAA*gttgaGGAGACTTAAGTCCAAAACC**TGG**	
Protoadjacent spacer motifs (PAMs) are in bold. Base pair spacers are lower case. Gix site or gix-related sites are in italics and dCas9 binding sites are underlined. * Chromosome_10 reporter contains two overlapping PAM sites and dCas9 binding sites on the 5' and 3' ends of the gix sites.		

[00157] Plasmids containing the recCas9 gene were constructed by PCR amplification of a gBlock encoding an evolved, hyperactivated Gin variant (Ginβ) (Gaj *et al.*, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic acids research 41, 3937-3946 (2013), the entire contents of which is hereby incorporated by reference)with the oligonucleotides 1GGS-rev-BamHI or 2GGS-rev-BamHI (using linker SEQ ID NO: 182) and Gin-for-NotI. PCR fragments were digested with BamHI and NotI, purified and ligated into a previously described expression vector (Addgene plasmid 43861) (see, *e.g.*, Fu *et al.*, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology 31, 822-826 (2013), the entire contents of which is hereby incorporated by reference) to produce subcloning vectors pGin-1GGS and pGIN-2GGS (using linker SEQ ID NO: 182). Oligonucleotides 1GGS-link-for-BamHI, 5GGS-link-for-BamHI (using linker SEQ ID NO: 701), or 8GGS-link-for-BamHI (using linker SEQ ID NO: 183) were used with Cas9-rev-FLAG-NLS-AgeI to construct PCR fragments encoding Cas9-FLAG-NLS with a 1, 5, or 8 GGS linker (see Table 3). For DNA sequences encoding the GGS amino acid linkers, see Table 7. PCR fragments and subcloning plasmids were digested with BamHI and AgeI and ligated to create plasmids pGinβ-2xGGS-dCas9-FLAG-NLS (using linker SEQ ID NO: 182), pGinβ-5xGGS-dCas9-FLAG-NLS (using linker SEQ ID NO: 701), and pGinβ-8xGGS-dCas9-FLAG-NLS (using linker SEQ ID NO: 183). For the DNA and amino acid sequence of the pGinβ-8xGGS-dCas9-FLAG-NLS (*i.e.*, recCas9), see below. The sequence encoding Ginβ is shown in bold; those encoding GGS linkers are shown in italics; those encoding dCas9 linkers are black; those encoding the FLAG tag and NLS are underlined and in lowercase, respectively.

ATGCTCATTGGCTACGTGCGCGTCTCAACTAACGACCAGAATACCGATCTTC
AGAGGAACGCACTGGTTTGTGCAGGCTGCGAACAGATTTTCGAGGACAAAC
TCAGCGGGACACGGACGGACAGACCTGGCCTCAAGCGAGCACTCAAGAGGC
TGCAGAAAGGAGACACTCTGGTGGTCTGGAAATTGGACCGCCTGGGTCGAA

GCATGAAGCATCTCATTTCTCTGGTTGGCGAACTGCGAGAAAGGGGGATCA
ACTTTCGAAGTCTGACGGATTCCATAGATACAAGCAGCCCCATGGGCCGGT
TCTTCTTCTACGTGATGGGTGCACTGGCTGAAATGGAAAGAGAACTCATTAT
AGAGCGAACCATGGCAGGGCTTGCGGCTGCCAGGAATAAAGGCAGGCGGTT
TGGAAGACCACCAAAG*GGTGGATCCGGAGGGTCCGGAGGTAGTGGCGGCAGCGG*
*TGGTTCAGGTGGCAGCGGAGGGTCAGGAGGCTCT*GATAAAAAGTATTCTATTGGTT
TAGCTATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGT
ACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAA
GAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGC
CTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTAC
TTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACC
GTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCA
TCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTA
TCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAAT
CTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTG
ATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAAC
CTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAA
GGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCA
CAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCAC
TAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATT
GCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAAT
TGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCG
CTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGC
CCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCG
AAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTAC
AAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTA
AAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGC
ATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGG
ATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTT
TCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGG
ATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTC
GATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAG
AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCA

CAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAAC
CCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGA
CCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTG
AATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACT
TGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAAC
GAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAG
ATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATA
AGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGC
GGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATT
TTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGA
CTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGA
CTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGC
ATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCAC
AAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAG
GGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGA
ACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAA
CGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAG
GAACTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAAT
CCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAA
CCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGA
ACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATA
ACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTAT
TAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACT
AGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGT
CAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAA
TTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTA
ATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGT
TTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGA
ACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAAT
TTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTA
ATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTC
GCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACT
GAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGT
GATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTC

GATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAA
AATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGC
GCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGA
AGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAA
AATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGA
ACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAG
AAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAG
CACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGA
GTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCAC
AGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTA
CCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAA
ACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATC
ACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGAC*GGTGGCT*
*CC*GATTATAAGGATGATGACGACAAG*GGAGGTTCC*ccaaagaagaaaaggaaggtcTGA
(SEQ ID NO: 184)

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQ
KGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYV
MGALAEMERELIIERTMAGLAAARNKGRRFGRPPK*GGSGGSGGSGGSGGSGGSG*
*GSGGS*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFD
SGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDK
KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFL
IEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ
LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ
YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS
RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY
FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE
ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDE
LVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTR

SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA
GFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQ
LFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGS*DYK
DDDDK*GGS*pkkkrkv Stop (SEQ ID NO: 185)

[00158] The Gin recombinase catalytic domain, which is amino acids 1-142 of SEQ ID NO: 185, is identical to the sequence of SEQ ID NO: 713. The dCas9 domain, in which is amino acids 167-1533 of SEQ ID NO: 185 is identical to the sequence of SEQ ID NO: 712.

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD
TLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAE
MERELIIERTMAGLAAARNKGRRFGRPPK (SEQ ID NO: 713)

DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL
NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY
KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTF
DNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVK
VMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDK

NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 712)

Table 7. DNA sequences encoding GGS linkers

GGS linkers	**SEQ ID NO:**	**DNA sequences for GGS linkers**	**SEQ ID NO:**
2XGGS	182	GGTGGTAGCGGTGGATCC	186
5XGGS	701	GGTGGATCCGGTGGTTCAGGTGGCAGCGGAGGGTCAG GAGGCTCT	187
8XGGS	183	GGTGGATCCGGAGGGTCCGGAGGTAGTGGCGGCAGC GGTGGTTCAGGTGGCAGCGGAGGGTCAGGAGGCTCT	188

[00159] For plasmid sequencing experiments, the AmpR gene in pGinβ-8xGGS-dCas9-FLAG-NLS (using linker SEQ ID NO: 183) was replaced with SpecR by golden gate cloning with PCR fragments. Esp3I sites were introduced into the pGinβ-8xGGS-dCas9-FLAG-NLS (using linker SEQ ID NO: 183) plasmid at sites flanking the AmpR gene by PCR with Esp3I-for-plasmid and Esp3I-rev-plasmid. The primers spec-Esp3I-for and spec-Esp3I-rev were used to amplify the SpecR marker as well as introduce Esp3I sites and Esp3I generated overhangs compatible with those generated by the Esp3I-cleaved plasmid PCR product. Golden gate assembly was performed on the two fragments following the protocol used to generate the reporter plasmids as described herein.

[00160] The pHU6-NT1 guide RNA expression vector was based on the previously described pFYF1328 (Fu *et al.*, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology 31, 822-826 (2013), the entire contents of which is hereby incorporated by reference) altered to target a region within the bacterial

luciferase gene LuxAB. Guide RNA expression vectors were created by PCR amplification of the entire vector with a universal primer R.pHU6.TSS(-1).univ and primers encoding unique guide RNA sequences (Table 1). A list of the guide RNA sequences is given in Table 8. These primers were phosphorylated with T4 polynucleotide kinase. The PCR reaction products and linear guide RNA expression vectors were blunt-end ligated and transformed. Guide RNA expression vectors used in initial optimizations, off target control guide RNA sequences and those targeting Chromosome 10 locus contained AmpR. All other plasmids described in this study contained specR to facilitate sequencing experiments. Spectinomycin resistance was initially introduced into guide RNA expression vectors via CPEC essentially as described (Quan *et al.*, Circular polymerase extension cloning of complex gene libraries and pathways. PloS one 4, e6441 (2009); and Hillson (2010), vol. 2015, pp. CPEC protocol; each of which is incorporated herein by reference) and guide RNA plasmids were then constructed by PCR amplification of the vector, as described above. Reactions were incubated overnight at 37 °C with 40 U of DpnI, purified and transformed. Fragments for CPEC were generated by PCR amplification of a guide RNA expression vector with oligonucleotides cpec-assembly-for-spec2 and cpec assembly-rev. The specR fragment was generated by PCR amplification of the SpecR gene via the oligonucleotides cpec-assembly-for-spec and cpec-assembly-rev-spec. pUC19 (ThermoFisher Scientific, Waltham, MA) was similarly modified.

Table 8. List of gRNA sequences

gRNA name	**gRNA-sequence**	**SEQ ID NO:**
on-target_gRNA	ACCTCTGTTTGGGAAAATTG	189
non-target_gRNA	gCACACTAGTTAGGGATAACA	190
Chromosome_10-54913298-54913376_gRNA-rev-5	gCCTCAGGGCCTGTGATGGGA	191
Chromosome_10-54913298-54913376_gRNA-rev-6	gCTCAGGGCCTGTGATGGGAG	192
Chromosome_10-54913298-54913376_gRNA-for-5	GGCCCATGACCCTTCTCCTC	193
Chromosome_10-54913298-54913376_gRNA-for-6	GCCCATGACCCTTCTCCTCT	194

Centromere_Chromosomes_1_5_19-gRNA-for	GACTTGAAACACTCTTTTTC	195
Centromere_Chromosomes_1_5_19-gRNA-rev	gAGTTGAAGACACACAACACA	196
Chromosome_5_155183064-155183141_(site 1)_gRNA-for	GGAACTCATGTGATTAACTG	197
Chromosome_5_155183064-155183141_(site 1)_gRNA-rev	gTCTACCTCTCATGAGCCGGT	198
Chromosome_5_169395198-169395274_(site 2)_gRNA-for	gTTTCCCGCAGGATGTGGGAT	199
Chromosome_5_169395198-169395274_(site 2)_gRNA-rev	gCCTGGGGATTTATGTTCTTA	200
Chromosome_12_62418577-62418652_gRNA-for	gAAATAGCACAATGAATGGAA	201
Chromosome_12_62418577-62418652_gRNA-rev	gACTTTTTGGGGGAGAGGGAG	202
Chromosome_13_102010574-102010650_(FGF14)_gRNA-for	GGAGACTTAAGTCCAAAACC	203
Chromosome_13_102010574-102010650_(FGF14)_gRNA-rev	gTCAGCTATGATCACTTCCCT	204
Off target-for (CLTA)	GCAGATGTAGTGTTTCCACA	205
Off target-rev(VEGF)	GGGTGGGGGGAGTTTGCTCC	206
Chromosome_12_**62098359-62098434_(FAM19A2)_gRNA-rev**	gATATCCGTTTATCAGTGTCA	207
Chromosome_12_**62098359-62098434_(FAM19A2)_gRNA-for**	gTTCCTAAGCTTGGGCTGCAG	208
Chromosome_12_**62112591-62112668_(FAM19A2)_gRNA-rev**	gCCTAAAAGTGACTGGGAGAA	209
Chromosome_12_**62112591-62112668_(FAM19A2)_gRNA-for**	gCACAGTCCCATATTTCTTGG	210

Cell culture and transfection

[00161] HEK293T cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) + GlutaMAX-I (4.5 g/L D glucose +110 mg/mL sodium pyruvate) supplemented with 10% fetal bovine serum (FBS, Life Technologies, Carlsbad, CA). Cells were cultured at 37 °C at 5% CO_2 in a humidified incubator.

[00162] Plasmid used for transfections were isolated from PureYield Plasmid Miniprep System (Promega, Madison, WI). The night before transfections, HEK293T cells were seeded at a density of $3x10^5$ cells per well in 48 well collagen-treated plates (Corning, Corning, NY). Transfections reactions were prepared in 25 µL of Opti-MEM (ThermoFisher Scientific, Waltham, MA). For each transfection, 45 ng of each guide RNA expression vector, 9 ng of reporter plasmid, 9 ng of piRFP670-N1 (Addgene Plasmid 45457), and 160 ng of recCas9 expression vector were mixed, combined with 0.8 µL lipofectamine 2000 in Opti-MEM (ThermoFisher Scientific, Waltham, MA) and added to individual wells.

Flow cytometry

[00163] After 60-72 hours post-transfection, cells were washed with phosphate buffered saline and harvested with 50 µL of 0.05 % trypsin-EDTA (Life Technologies, Carlsbad, CA) at 37 °C for 5-10 minutes. Cells were diluted in 250 µL culture media and run on a BD Fortessa analyzer. iRFP fluorescence was excited using a 635 nm laser and emission was collected using a 670/30 band pass filter. EGFP was excited using a 488 nM laser and emission fluorescence acquired with a 505 long pass and 530/30 band pass filters. Data was analyzed on FlowJo Software, gated for live and transfected events (expressing iRFP). Positive GFP-expressing cells were measured as a percentage of transfected cells gated from at least 6,000 live events. For optimization experiments, assay background was determined by measuring the percentage of transfected cells producing eGFP upon cotransfection with reporter plasmid and pUC, without recCas9 or guide RNA expression vectors. This background was then subtracted from percentage of eGFP-positive cells observed when the reporter plasmid was cotransfected with recCas9 and the on-target or non-target guide RNA expression vectors.

Identification of genomic target sites

[00164] Searching for appropriate target sites was done using Bioconductor, an open-source bioinformatics package using the R statistical programming (Fu *et al.*, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology 31, 822-826 (2013), the entire contents of which is hereby incorporated by reference). The latest release (GRCh38) of the human reference genome published by the Genome Reference Consortium was used to search for sites that matched both the PAM requirement of Cas9 and the evolved gix sequence as described in the text. With the genome loaded into R, each search pattern was represented as a Biostring, a container in R that allowed for string matching and manipulation Scanning both strands of DNA for the entire genome, using the stated parameters, reveals approximately 450 potential targets in the human genome when searching using the GRCh38 reference assembly (Table 9).

Table 9. recCas9 genomic targets identified *in silico*

Chr.	Start	End	Sequence	Pattern ID	SEQ ID NO:
chr1	34169027	34169103	CCTTTAGTGAAAAGTAGACAGCTCTGAATAT GAAAGGTAGGTTTTCATTTCTGGGAAAGAGA CGCCAAGTGATGTGG	2	211
chr1	51006703	51006780	CCTCCAATAAATATGGGACTATGTGGAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACGGGAAGAATGG	1	212
chr1	89229373	89229450	CCATTCTGCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTAGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	213
chr1	115638077	115638154	CCATTCTCCCCGTCACTTTCAGGTACAACAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	214
chr1	122552402	122552478	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTGTTGTGG	2	215
chr1	122609874	122609950	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCATACTTGAA ACACTCTTTTTGTGG	2	216
chr1	122668677	122668753	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	217
chr1	123422419	123422495	CCTTGTGTTGTGTTTATTCAACTCACAGAGTT AAACGATCCTTTACACAGAGCAGACTTGAA ATACTCTTTTTGTGG	2	218
chr1	123648614	123648690	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCATACTTGAA ACACTCTTTTTGTGG	2	219
chr1	123806335	123806411	CCTTGTATTGTGAGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	220
chr1	124078228	124078304	CCTTGTGTTGTGTGTCTTCAACTCACAGAGTT AAACGATGCTTTACACAGAGTAGACTTGAA ACACTCTTTTTCTGG	2	221
chr1	124231074	124231150	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGTA ACACTCTTTTTGTGG	2	222

chr1	124232435	124232511	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACGTGA AACACTCTTTTTGTGG	2	223
chr1	124344781	124344857	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	224
chr1	124435716	124435792	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGGAGACTTGTA ACACTCTTTTTGTGG	2	225
chr1	158677186	158677262	CCTGAGGTTTTCCAGGTTTTAAAAGGAAACC TAAAGGTAGGTTTAGCATTAAGTGTCTTGAA GTTTATTTTAAAAGG	2	226
chr1	167629479	167629554	CCAAAATTCCCACAAAACCGAATGCATCAGT CAAAGCAAGGTTTGAAGAAAAGATTTACCA CTTCAGGGAGCTTGG	4	227
chr1	167783428	167783504	CCTTTTCTGGATATCGTTGATGCTCTGTATGC AAAAGGTAGGTTTTTGGGTTATGTTGTTAAA CAGTGATTGAATGG	3	228
chr1	169409367	169409444	CCTCCAAGAAATATGGAACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGAGAGAATGG	1	229
chr1	174145346	174145423	CCTCCAAGAAATATGGGACTATGTGAGAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	230
chr1	183750168	183750245	CCATTCTCCCCATCGCTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTT CCATATTCTTTGGAGG	1	231
chr1	200801540	200801617	CCATTCTCCCCATCACTTTCAGGTGTACCGA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	232
chr1	207589936	207590013	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACGGGGAGAATGG	1	233
chr1	209768370	209768445	CCTTCAGGGCAGAAACAGCTCTACTAGCAG AGAAAGCAAGCTTTCAATATTGTGCAATACA AAAACGAGAGCAGGG	4	234
chr1	218652378	218652455	CCATTCTCCTCATCTCCTTCTGGTACTCCAAT CAAACGTAGGTTTGGTCTTTTCTCATAGTCTC ATATTTCTTGGAGG	1	235
chr1	222147250	222147327	CCTCCAAGACATATAGGACTATGTGAAAATA CCAAACCTACGTTTGATTGGTGTACCTGAAA GTGACAGGGAGTATGG	1	236
chr1	245870710	245870785	CCTGCCAGATACCAGTAGTCACTGTGAATTA CAAAGCTACGTTTCTTCCATAGGGAAAGTTT GGAGTCCAGCCAGG	4	237
chr2	2376037	2376114	CCATTCTCCCTGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	238
chr2	4119629	4119706	CCATTCTCCCCACCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGTAGG	1	239
chr2	4909047	4909124	CCTAACCAGAAACTAACTAATAGATATGGG CAGAAAGCATCCTTTCACTTTTGTTCTGGGA GAGGGAAGAAGCAAAGG	1	240
chr2	28984877	28984953	CCATTTTGGGGAGGCCTTGATGGGAAGCTGG AAAAGGAAGCTTTCCTCCCAGTCCTGCTGAA GGCCTTGCCAGCTGG	2	241
chr2	31755833	31755910	CCTCCAAGAAACACAGGACTATGTGAAAAG ATCAAACCTACGTTTGATTGGTGTTCCTGAA AGTGATGGGGAGAATGG	1	242
chr2	39829583	39829660	CCATTCTCTTCATGACTTTCAGGTACACCATT GAAACGTAGGTTTGGTCTTTTCACATTGTCC CATATTTCTTGGAGG	1	243
chr2	60205947	60206024	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCGTATTTCTTGGTGG	1	244

chr2	79082362	79082439	CCATTCTCCCTGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGGGG	1	245
chr2	79082362	79082438	CCATTCTCCCTGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGGG	3	246
chr2	108430915	108430992	CCTCCAAGAAATATGAGATTATATGAAAAG ACCAAACCTACGTTTGATTGGTGTACTTTAA AGTGACGGGGAGAATGG	1	247
chr2	115893685	115893762	CCATTCTCCCCGTCATTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCAAATTTCTTGGAGG	1	248
chr2	119620068	119620145	CCCCCAAGAAATGTGGGACTATATGAAAAG ACCAAACCTACGTTTGACTGGTGTACCTAAA AGTGATGGGGAGAATGG	1	249
chr2	119620069	119620145	CCCCAAGAAATGTGGGACTATATGAAAAGA CCAAACCTACGTTTGACTGGTGTACCTAAAA GTGATGGGGAGAATGG	2	250
chr2	128495068	128495144	CCCATTGGTGCTGACCAGATGGTGAAGGAG GCAAAGGTTGCTTTGAATGACTGTGCTCTGG GGTGAGCCAGGCCTGG	2	251
chr2	133133559	133133634	CCCTTTACAGAGGTGAGCTTTGTTATTAGTA AAAAGGTAGGTTTCCCTGTTTTTCTGAAGAA AAGCTGTGAGTGGG	4	252
chr2	134174983	134175060	CCACTGCCCATTGACAGAGTGGCGAGGTGG GTGAAACCTTGCTTTCCTCCTGGCCCATGGG CAGGGTGGGGCTGTGGG	1	253
chr2	134174983	134175059	CCACTGCCCATTGACAGAGTGGCGAGGTGG GTGAAACCTTGCTTTCCTCCTGGCCCATGGG CAGGGTGGGGCTGTGG	3	254
chr2	138069945	138070022	CCATTCTCCCTGTCACTTTTAGATACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATGTTTCTTGGAGG	1	255
chr2	138797420	138797496	CCTCCAAGAAATATCAACTGTGTGAAAAGA CGAAACCTACGTTTGATTAATGTACCTGAAA GTGACAGGGAGAATGG	2	256
chr2	145212434	145212511	CCATTCTCCCATTAACTTTCAAGTACACCAA TCAAAGGTAGGTTTGGTGTTTTCCCATAGTC CCGTATTTCTTGGAGG	1	257
chr2	147837842	147837919	CCTTTTCATCATGCCCCTTTCACTTTAAGGTG AAAACCTTGCTTTACATGTCAGAGAAAAGA AGAGCCCTCAGCTGGG	1	258
chr2	147837842	147837918	CCTTTTCATCATGCCCCTTTCACTTTAAGGTG AAAACCTTGCTTTACATGTCAGAGAAAAGA AGAGCCCTCAGCTGG	3	259
chr2	154152540	154152617	CCATTCACCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	260
chr2	157705943	157706019	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATGGTGTACCCGAAA GTGACAGGGAGAATGG	3	261
chr2	158361152	158361229	CCACCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATAGGTATACCTGAA AGTGACAGGGAGAATGG	1	262
chr2	161461006	161461083	CCATTCTCCCCATCACTTTCAGGTGCACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	263
chr2	179077376	179077453	CCCTCAAGAAATATGAGACTATGTGAAAAG ACCAAACCTACGTTTGACTGGTATACCTGAA AGTGACAGGGAGAATGG	1	264
chr2	179077377	179077453	CCTCAAGAAATATGAGACTATGTGAAAAGA CCAAACCTACGTTTGACTGGTATACCTGAAA GTGACAGGGAGAATGG	2	265
chr2	181090699	181090776	CCTCCAACAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACGGGGATAATGG	1	266

chr2	182331957	182332034	CCATTCTCTCCCTCACTTTCAAGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCT TATATTTCTTGGCGG	1	267
chr2	183620562	183620638	CCATTCTCCCTGTCACTGTCAGTACACCAAT CAAACGTAGGTTTGGTCTCTTCACATAGTCC CATATTTCTTGGAGG	2	268
chr2	207345927	207346003	CCTCCAAGAAATATGGGACTATGTGAACAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGATGGCAGAATGG	3	269
chr2	216652047	216652123	CCACCATGCCTGGCCACCACACATTTTTTTCT AAAGCTTGGTTTTGGCCACAGTGAGAGTTTC TTGGGCTGTCAGGG	2	270
chr2	216652047	216652122	CCACCATGCCTGGCCACCACACATTTTTTTCT AAAGCTTGGTTTTGGCCACAGTGAGAGTTTC TTGGGCTGTCAGG	4	271
chr2	223780040	223780116	CCCACTAGGTGGCGATATCTGAGGGTCCAAT GAAACCATGCTTTTTACTCAGATCTTCCACT AACCACCTCCCCCGG	2	272
chr2	224486595	224486672	CCTCTAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGACTGGTGTACCTGAA AGTGACGGGGAGAATGG	1	273
chr2	230526902	230526979	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTAGTGTACCTGAA AGTGACGGGGAGAATGG	1	274
chr2	232036127	232036204	CCATTCTCCCTGTCACTTTCAGGTACATCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	275
chr3	4072812	4072889	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGACTGGTGTACCTGAA AGGGATGGGGAGAATGG	1	276
chr3	9261677	9261754	CCCCCAAGAAATATGAGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	277
chr3	9261678	9261754	CCCCAAGAAATATGAGACTATGTGAAAAGA CCAAACCTACGTTTGATTGGTGTACCTGAAA GTGACAGGGAGAATGG	2	278
chr3	16732146	16732223	CCTCTAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTAACTGAA AGTGACAGGGAGAATGG	1	279
chr3	17450712	17450789	CCTCCAAGAAATATGCGCCTATGTGAAAAG ACCAAACCTACGTTTGATTGGTATACCTGAA AGTGATGGAGAGAATGG	1	280
chr3	21559769	21559846	CCATTCTCCCTGTCACTTTGAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATATTC GCATATTTCTTGGAGG	1	281
chr3	23416658	23416735	CCATTCTCCCCGTCACTTTCAGGTACACCAA CCAAACGTTGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	282
chr3	29984019	29984096	CCATTCTCCCTGTCACTTTCCAGTACACCAGT CAAACGTAGGTTTGGTCTTTTCACATACTCC CATATTTCTTGGAGG	1	283
chr3	38269551	38269627	CCTGGCCTAATTTTTAATTCTTAGTTTGACTT AAACCTTGCTTTTAGTGTGATGGCGACAAAA GCTGAGCTGAAAGG	2	284
chr3	40515213	40515288	CCAGTGCTTTTTGGTTTTAAAGGCAAGCCTC CAAACCTTCCTTTCTCCTGGATGCTGTGGTG GTTGCCATGCATGG	4	285
chr3	49233612	49233687	CCCAACTCCTGCGAGAAGTAGCTCACCATGA CAAAGCTACCTTTGCTTTTATCGTTTTGCAAA ACAAAAAAGGGGG	4	286
chr3	66292894	66292971	CCATTCTCCCCGTCACTTTGAGGTGTGCCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CTATATTTCTTGGAGG	1	287
chr3	67541493	67541570	CCTCCAAAAAATATGGGACTACGTAAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA ACTGACAGGGAGAATGG	1	288

chr3	82273011	82273088	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTT CCATATTTCTTGGAGG	1	289
chr3	98683349	98683426	CCTACAAGATATATGGGACTATGTGAAAAG ACCAAACCTACGTTTTACTGGTGTGCCTGAA ACTGACGGGGAGAATGG	1	290
chr3	101923653	101923730	CCATTCTCTCTGTCACTTTCAGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	291
chr3	114533467	114533544	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTCATTGGTGTACCTGAA AGTGATAGGGAGAATGG	1	292
chr3	132607602	132607679	CCTCCAAAAAATATGGGATGATGTGAAAAG ACCAAACCTAGGTTTGACTGGTGTACCTGAA AATGATGGGGAGAATGG	1	293
chr3	137545176	137545253	CCTCCAAGAAATATGAGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	294
chr3	137655679	137655756	CCTCCAAGAAATATGGGACTACGTGAAAAG ATCAAACCTACGTTTGATTGTTGTACCTGAA AGTGATGGGGAGAATGG	1	295
chr3	137662040	137662117	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGTTGTACCTGAA AGTGATGGGGAGAATGG	1	296
chr3	142133796	142133873	CCTCAAAAGTGTTCTGGTTTTGTTTTGTTTTT TAAACCATGGTTTTACCTCTGGCTTAGTGGG ACTAAAAATAGGAGG	1	297
chr3	146726949	146727026	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGACTGGTGTACCTGAA AGTGATGGGGAAAATGG	1	298
chr3	152421096	152421173	CCTCCAAGAAATATGGGACTGTGTGTAAAG ACCAAACCTACGTTTGATTGGTGTACCTCAA AGTGATGGGGAGAATGG	1	299
chr3	170620247	170620324	CCATTCTCCCCATCACATTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	300
chr3	181166873	181166949	CCCCTGGAAAAGTTGGAGCATCACAGGAAA AGCAAACCAACCTTTTTTCTCCCCTAGGTAA ACTGGGGAGCCAGGGG	3	301
chr3	181166874	181166949	CCCTGGAAAAGTTGGAGCATCACAGGAAAA GCAAACCAACCTTTTTTCTCCCCTAGGTAAA CTGGGGAGCCAGGGG	4	302
chr4	6604233	6604309	CCTTCCCCAGTTGCAGCAGACAAGAGTCTCG AAAAGCTTGCTTTGGTTGCTGCAGTGGATGG GTTGGTAGGCACAGG	2	303
chr4	6626269	6626344	CCCCCACCTCCCAAGCTGCTGGCTTCTCGAA TAAAGCTACCTTTCCTTTTACCAAAACTTGTC TCTCGAATGTCGG	4	304
chr4	8155396	8155472	CCTTGGCCCTGGACAGCTGCTTTTCCTTCCCT AAACCTTGGTTTCCCCCTTTGTGCAGGTGGG TGGGTTTGGGCTGG	2	305
chr4	10386803	10386880	CCTCTTCTAGTGAACCCATGGGGTTACCAAG GGAAAGCAACCTTTTGATAAATATTCCCATC TTTTTATGTTGTCTGG	1	306
chr4	20701579	20701656	CCACTTGAAAGGGTTACCAAGGATAAGATTT TTAAAGCTTGCTTTCACAAACAACTCATGCT CCAGGCTTGTCAGTGG	1	307
chr4	29594286	29594363	CCTTTCTCCCCATCACTTTCAGGTACACCAAT CAAACGTAGGTTTGATCTTTTCACATAGTCC CATATTTCTTGGAGG	1	308
chr4	53668422	53668499	CCATTCTCCCCATCAATTTCAGTTACACCAA TGAAACGTAGGTTTGGCCTTTTCACATAGTC CCATATTTCTTAGAGG	1	309
chr4	74914802	74914879	CCATTCTCCCTGTCACTCTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCATATAGTC CCATATTTCTTGGAGG	1	310

chr4	75332783	75332859	CCTCCAAGAAAATTGGGACTATGTGAAAAA ACCAAACCTACGTTTGATTGATGTACCTGAA AGTGACAGGAGAATGG	3	311
chr4	88123643	88123720	CCTTCAAGAAATATGGGACTATGTGAAAGG ACAAAACCTACGTTTTATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	312
chr4	89567192	89567269	CCATTCTCCCCATCACTTTCAGGTACGCTAA TCAAACGTAGGTTTGATCTTTTCACATAGTC TTATATTTCTTGGAGG	1	313
chr4	93556577	93556654	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGACTGGTGTACCTCAA TGTGACAGGGAGAATGG	1	314
chr4	100266379	100266456	CCATTCTCCCTGTCACTTTTAGGTACACCAAT CAAACGTACGTTTGGTCTTTTCACATAGACC CATATTTCTTGGAGG	1	315
chr4	103486234	103486311	CCTTCAAGAAATATGGGACTGTGTGAAAAG ACCAAAGCTAGGTTTGATTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	316
chr4	105923129	105923204	CCTACTATTCACAGAGTAATGCAGTTTGCTG AAAAGGTTGGTTTTTGCTGACCTCTGAGAGC TCACATTACAGTGG	4	317
chr4	106874711	106874788	CCATTCTCTCTGTCACTTTCTGGTACACCAAT CAAACGTAGGTTTGCTCTTTTCACATAATCC CATATTTATTGAAGG	1	318
chr4	115805791	115805867	CCATAACATGTATTTGCTGGTGCTAGACTCT CCAAAGCTAGGTTTCTTTCTACAACAATGGC TGGAAGTCTTCTTGG	3	319
chr4	122033277	122033354	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTCTCACACAGTC CCATATTTCTTGGAGG	1	320
chr4	129125132	129125209	CCATTCTTCCCATTACTTTCAGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CACATTTCTTGGAGG	1	321
chr4	135472562	135472639	CCATTCTCCCCCTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATTGTCC CATATTTCTTGGAGG	1	322
chr4	138507099	138507176	CCATTCTCCCCAGCACTTACAGGTACACCAA TCAAACGTAGGTTTGGTCATTTCACATAGTC CCATATTTCTTGGAGG	1	323
chr4	144249093	144249170	CCATTCTCCCTGTCACTTTCAGGTACAGCAA TCAAACGTAGGTTTGGTCTTTTCACATGGTC CCATATTTCTTGGAGG	1	324
chr4	144436406	144436483	CCTCCAAGAAATATGAGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACGGGGAAGATGG	1	325
chr4	154110259	154110336	CCTCCAAGAAATATGAGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	326
chr4	154893438	154893515	CCTCCAAGAGATATGAGACTATGTAAATAG ACCAAACCTACCTTTGATTGGTGTACGTGAA AGTGACAGGAAGAATGG	1	327
chr4	161116854	161116931	CCATTCTCCCCATCACTTTCAGGTACACCAA CCAAACGTAGGTTTGGTCTTTTCACATAGTC TCATATTTCTTGGAGG	1	328
chr4	165140748	165140823	CCTCCATTGACTACTCCTTATCATTGGCTAG AAAACCTACCTTTCAACCAGTTTCTAAGGCC AAGAAACTTGGAGG	4	329
chr4	181928508	181928585	CCACCAAGAAATATGGGACTACGTGAAAAG ACCAAACCTACGTTTGATGGGTGTGCCTGAA AGTGACGGGAAGAATGG	1	330
chr4	187521958	187522035	CCTCCAAGAAATAAGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA GGTGACAGGGAGAATGG	1	331
chr5	12675639	12675715	CCAAAGGGCCTTTGTGATTCTACTTTGTAAT ATAAAGGATGGTTTCTTACTACGGTTGGTGT CCTTGCAGGAGTGGG	3	332

chr5	29271804	29271881	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	333
chr5	35352660	35352737	CCATTCTCCCCGTTACTTTCAGGTACACCAA TAAAACCTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	334
chr5	38723235	38723310	CCCATATCTCTGGCAAGGGCAGCTCTCTGGC TAAACCAAGCTTTCCTGTAGAGCTTGAGTTC CAAGGCAGCGTTGG	4	335
chr5	47358339	47358415	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTGTTGTGG	2	336
chr5	47415811	47415887	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCATACTTGAA ACACTCTTTTTGTGG	2	337
chr5	47474614	47474690	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	338
chr5	48228356	48228432	CCTTGTGTTGTGTTTATTCAACTCACAGAGTT AAACGATCCTTTACACAGAGCAGACTTGAA ATACTCTTTTTGTGG	2	339
chr5	48454551	48454627	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCATACTTGAA ACACTCTTTTTGTGG	2	340
chr5	48612272	48612348	CCTTGTATTGTGAGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	341
chr5	48884165	48884241	CCTTGTGTTGTGTGTCTTCAACTCACAGAGTT AAACGATGCTTTACACAGAGTAGACTTGAA ACACTCTTTTTCTGG	2	342
chr5	49037011	49037087	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGTA ACACTCTTTTTGTGG	2	343
chr5	49038372	49038448	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACGTGA AACACTCTTTTTGTGG	2	344
chr5	49150718	49150794	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	345
chr5	49241653	49241729	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGGAGACTTGTA ACACTCTTTTTGTGG	2	346
chr5	88582714	88582790	CCTTTTCATAAGAAGAAAATCGACTCATCAT TGAAACCAAGCTTTGGTACAATTTCATTGAT GTTTCCAGAAGCAGG	3	347
chr5	93497156	93497231	CCCATAGACTATGATAGAAACAAAATAACC CAAAAGCTAGCTTTCTGATTGAGTTTCCATA AATGCAATGTGAAGG	4	348
chr5	94295029	94295105	CCATTCACTTGTCACTTTCTGGTACACCAATC AAACGTAGGTTTGGTCTTTTCACATAGTCTC ATATTTCTTGGAGG	2	349
chr5	94956746	94956823	CCTCCAAGAAATATGGGACTCTGTAAAGAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGAAGGGGAGAATGG	1	350
chr5	106003488	106003565	CCATTCTCCCCGTCATTTTCAGGTACACCAA TCAAACCTAGGTTTGGTCTTTTTACATAGTCC CATATTTCTTGGAGG	1	351
chr5	118727905	118727982	CCTCCACGAAACATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	352
chr5	132156032	132156109	CCAATTTCCCCCTCACTTTCAGATACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTT CCATATTTCCTGGAGG	1	353
chr5	152037951	152038028	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATATTCC CATATGTCTTGGAGG	1	354

chr5	155183064	155183141	CCCACCGGCTCATGAGAGGTAGAGCTAAGG TCCAAACCTAGGTTTATCTGAGACCGGAACT CATGTGATTAACTGTGG	1	355
chr5	155183065	155183141	CCACCGGCTCATGAGAGGTAGAGCTAAGGT CCAAACCTAGGTTTATCTGAGACCGGAACTC ATGTGATTAACTGTGG	2	356
chr5	163148211	163148288	CCTTCAAGAAATATGGGACTATGTGAAGAG ACCAAACCTACGTTTGATTGGTGTAGCCAAA AGTGATGGGGAAAATGG	1	357
chr5	165889537	165889614	CCTCAGATTAGATTTACTTGCAAAGAGACAT TTAAAGGATCGTTTTGATACTATTTTGAAAG TACTATACAAAGATGG	1	358
chr5	169395198	169395274	CCTTAAGAACATAAATCCCCAGGAATTCACA GAAACCTTGGTTTGAGCTTTGGATTTCCCGC AGGATGTGGGATAGG	2	359
chr5	171021380	171021457	CCATTCTCTCTGTCACTTTCAGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCTCATAGTCC CATATTTCTTGGAGG	1	360
chr5	173059898	173059973	CCATTTACCATCATTCTCTGTCATGGCAGGT GAAAGCAAGCTTTTATATAGACAATGTTCTA CTTAGTTTACAGGG	4	361
chr5	174102359	174102435	CCCAAAGTTAATTTTACTCTTTTTCTGAATCA AAAGGAACCTTTCCTCCATGAGAAGAATCCT GCCATATTTCTAGG	2	362
chr5	180927811	180927888	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGCTATACATGAA AGTGACGGGGAGAATGG	1	363
chr6	1752363	1752440	CCTTCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACCTTTGATTGGTGTACCTGAA AGTGATGGGAAGAATGG	1	364
chr6	20595279	20595356	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATAGTTCTTGGAGG	1	365
chr6	23431370	23431447	CCATTCTCCCCGTCACTTTCAGGGACAACAA TCAAACGTAGGTTTGGCCTTTGCACATAGTC TTATATTTCTTGGAGG	1	366
chr6	29190624	29190701	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	367
chr6	61533266	61533343	CCTCCAAAAAATATGGGACTATGTGAGAAG ACCAAACCTACGTTTTATTAGTGTACCTCAA AGTGACAGGGAGGATGG	1	368
chr6	101052764	101052841	CCATTCTCCCCATCACTTTCAGGTACACCAA TGAAACGTAGGTTTGGCCTTTTCACATAGTT TCATATTTCTTGGAGG	1	369
chr6	117176355	117176432	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	370
chr6	117747073	117747149	CCTACAAGAAATATGGAACTTGTAAAAAGA CCAAACCTACGTTTGATTGGTGTACCTGAAA GTGACGGGGAGAATGG	2	371
chr6	118422508	118422585	CCTCCAAGAAATATGGGACAATGTGAAAAG GCCAAAGCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	372
chr6	122035019	122035096	CCTTTCAAACTTAGAGGTAAACAAAAGTCCT GAAAACCTAGGTTTGACCATAAGTTGGGACC ATACGAGCATAGAAGG	1	373
chr6	134445210	134445287	CCAAAAATAAAAAAAAAATTGACTTATAAGT AAGAAAGGTTCGTTTTCTCACATTCAGAAAG AGAACCCACATGTTGGG	1	374
chr6	134445210	134445286	CCAAAAATAAAAAAAAAATTGACTTATAAGT AAGAAAGGTTCGTTTTCTCACATTCAGAAAG AGAACCCACATGTTGG	3	375
chr6	135154944	135155021	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	376

chr6	137889995	137890072	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTTGGTTTAGTCTATTCACATAGTC CCATATTTCTTGGAGG	1	377
chr6	143993904	143993981	CCGAAAAGAATAAGACTATCAGCTGAAGTC TTAAAACGATCCTTTGGCCCCCAGTACTCTA TATGCAGGATAGAAAGG	1	378
chr6	152610473	152610549	CCTACAAAAATAGGGGACTATGTGATAAGA CCAAACCTACGTTTGATTGGTGTACCTGAAA GTGATGGGGAGAATGG	2	379
chr6	160372604	160372681	CCATTCTACCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGCCTTTTCATATAGTC TCATATTTCTTGGAGG	1	380
chr6	169352478	169352555	CCATTCTCCCCATCACTTTCTGGTATACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTAGAGG	1	381
chr6_GL000 251v2_alt	677196	677273	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	382
chr6_GL000 252v2_alt	456242	456319	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	383
chr6_GL000 253v2_alt	456202	456279	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	384
chr6_GL000 254v2_alt	456371	456448	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	385
chr6_GL000 255v2_alt	456225	456302	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	386
chr6_GL000 256v2_alt	500011	500088	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	387
chr7	5256551	5256627	CCACCACACCCAGCCTTATGGGATGGTTTTC AAAAGCATCCTTTTTTAGAAGTGGATTCTGA TATATAATCGGATGG	2	388
chr7	7392583	7392660	CCATTCTCAATGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	389
chr7	8737741	8737818	CCATTCTCTCTGTCACTTTCAGGTACACCAGT CAAAGGTAGGTTTGTTTTATTCACACGTTCA CATATTTCTTGGAGG	1	390
chr7	11352226	11352303	CCATTCGCCCCATCACTTTCAGGTACACTAG TAAAACGTAGGTTTGGTCTTTTCACATAGTT CCATATTTCTTGGAGG	1	391
chr7	15519145	15519222	CCTCCAAGAAATATGGGACTATGTGAAGAG ATCAAACCTAGGTTTGATTGTTGTACCTGAA AGTGATAAGAAGAATGG	1	392
chr7	19228341	19228418	CCTCCAATAAATATGGGGCTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	393
chr7	23778445	23778522	CCCTTTTCCCTGTCACTTTCAGGTACACCAGT CAAACGTAGGTTTGGTCTTTTCACATAGTCG AATATTTCTTCAAGG	1	394
chr7	23778446	23778522	CCTTTTCCCTGTCACTTTCAGGTACACCAGTC AAACGTAGGTTTGGTCTTTTCACATAGTCGA ATATTTCTTCAAGG	2	395
chr7	26769065	26769142	CCATTCTCCCTGTCACTTTCAGGTACACTAAT CAAACGTAGGTTTGGTGTATTCACACAGTCC CATATTTCTTGGAGG	1	396
chr7	42864035	42864112	CCATTCTTCCTGTCACTTTCAGGTATACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATGTTTCTTGGAGG	1	397
chr7	46498923	46499000	CCTCCAAGAAATATGAGACTATATGAAAAT ACCAAACCTACGTTTGATTGGTGTACCTGAA AGAGACAGGGAGAATGG	1	398

chr7	51535360	51535437	CCATTCTCCCTATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCATGTAGTC CCATATTTCTTGGAGG	1	399
chr7	51927106	51927183	CCATTCTGCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	400
chr7	56976942	56977018	CCGTCCGATTATATATCAGAATCTACTTCTA AAAAAGGATGCTTTTGAAAACCATCCCATAA GGCTGGGTGTGGTGG	3	401
chr7	80021598	80021675	CCTACAAGGAATATAGGACTATGTGAAAAT ACCAAACCTACGTTTCACTGCTGTACCTGAA GGTGACAGGGAGAATGG	1	402
chr7	89673853	89673930	CCATTCTCCCCATCATTTCCAGGTAAACCAA TCAAAGGTAGGTTTGGTCATTTCACATAGTC CCATATTTCTTGGAGG	1	403
chr7	103404790	103404867	CCATTCTCCCCGTCACTTTCAGGTACACCAG TCAAACGTAGGTTTGGTCTTTTCACACAGTC CCATATTTCCTGGAGG	1	404
chr7	113053651	113053728	CCATTCTCCCCATCACTTTCAGGTACAGCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	405
chr7	125765204	125765279	CCACTACAGATTCTTGGGTCAAGATGTGTGC AAAAGGATGCTTTAGGGTGATGGATATGAG TGGGATGAAATGAGG	4	406
chr7	128042158	128042234	CCTGAAAAAAAACCCTGCCAGCCAGCAACT CTGAAAGGATGCTTTGTGTGAGTGAGCAGTG TCTGAGATGGACAGGG	3	407
chr7	130637332	130637409	CCATTCTCCCCATCACTTTCAGGTACGCCAA TCAAACGTAGGTTTGGTCTTTTGACATAGTC CCATATTTCTTGGAGG	1	408
chr7	136983050	136983127	CCGTTCTCCCCATCACTTTTAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC TCATATTTCTTGGAGG	1	409
chr7	143579507	143579584	CCATTCTCCTGGTCACTTTCAGGTATACCAA TCAAACGTAGGTTTGGTCTTTTCATGTAGTC CCATATTTCTTGGAGG	1	410
chr7	143749881	143749958	CCTCCAAGAAATATGGGACTACATGAAAAG ACCAAACCTACGTTTGATTGGTATACCTGAA AGTGACCAGGAGAATGG	1	411
chr8	2338364	2338441	CCTCCAAGAACTATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACGGGGAGAATGG	1	412
chr8	2383289	2383366	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATAGTTCTTGGAGG	1	413
chr8	8414568	8414645	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACAGAGTC CCATATTTCTTGGAGG	1	414
chr8	24163142	24163219	CCATTCTCCCCGTCACTTTCATGTACACCAA GCAAACGTAGGTTTGATCTTTCCACATAGTC CCGTGTTTCTTGGAGG	1	415
chr8	34299051	34299128	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACTTGAA AGTGACAGGGAGAATGG	1	416
chr8	40965485	40965562	CCTCCAAGAAATATGGGACTATGTGAAAAG ACAAAACCTACGTTTCACTGGTGTACCTGAA AGTGACAGGGAGGATGG	1	417
chr8	48371659	48371735	CCCCCACCTTTTAAAAACATGCATACATACG GAAACGTTGCTTTCTGCACGATTTCATTTTA ATGGAACAGAACAGG	2	418
chr8	82534960	82535037	CCATTTCCCCTGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTA TCATATTTCTTGGAGG	1	419
chr8	109217624	109217700	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTGGAGG	3	420

chr8	134790285	134790361	CCTTTTGTTAAAGTAATAGAATTCTGCTTCTTAAAGGAACCTTTCAGGCAAGATGGTGGTTAGAGCACCTAAATGGG	2	421
chr8	134790285	134790360	CCTTTTGTTAAAGTAATAGAATTCTGCTTCTTAAAGGAACCTTTCAGGCAAGATGGTGGTTAGAGCACCTAAATGG	4	422
chr8_KI270821v1_alt	519635	519712	CCTCCAAGAACTATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACGGGGAGAATGG	1	423
chr8_KI270821v1_alt	564557	564634	CCATTCTCCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGCCTTTTCACATAGTCCCATAGTTCTTGGAGG	1	424
chr9	14951207	14951283	CCTCCAAGAAATATGGGACTGGTGAAAAGACCAAACCTACGTTTGACTGGTGTACCTGAAAGTGACGGGGAGACTGG	2	425
chr9	23249218	23249295	CCTCCAAGAAACATGGGAATGTGTGAAAAGACCAAACCTACGTTTGATTGGCGTACCTGAAAGTGACGGGGAGTATGG	1	426
chr9	26278896	26278973	CCTCCAAGAAATATGGGACTGTGTGAAAAGACCAAACCTACGTTTGATTGGTATACCTGAAAGTGACAGAGAGAATGG	1	427
chr9	27323237	27323314	CCATTCTCCCCTTCACTATCAGGTACACCAATCAAACGTAGGTTTAGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	428
chr9	31517993	31518070	CCATTCTCCCCGTCACTTTCAGATACACCAGTCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	429
chr9	39694860	39694937	CCATCTTACTTTGTACTACACTGTTCTTTAGAGAAAGCTTCCTTTTGGAGACCAACCAGGACTCCTTAGAAGCAGAGG	1	430
chr9	42451132	42451209	CCATCTTACTTTGTACTACACTGTTCTTTAGAGAAAGCTTCCTTTTGGAGACCAACCAGGACTCCTTAGAAGCAGAGG	1	431
chr9	60776573	60776650	CCTCTGCTTCTAAGGAGTCCTGGTTGGTCTCCAAAAGGAAGCTTTCTCTAAAGAACAGTGTAGTACAAAGTAAGATGG	1	432
chr9	62647482	62647559	CCTCTGCTTCTAAGGAGTCCTGGTTGGTCTCCAAAAGGAAGCTTTCTCTAAAGAACAGTGTAGTACAAAGTAAGATGG	1	433
chr9	66682030	66682107	CCTCTGCTTCTAAGGAGTCCTGGTTGGTCTCCAAAAGGAAGCTTTCTCTAAAGAACAGTGTAGTACAAAGTAAGATGG	1	434
chr9	82264427	82264503	CCACCACTGTGCCTGGCCATTTTCACTATTCTTAAAGGAAGCTTTGGTTTACAAAGGTTTGCTACTGTACTTCCAGG	3	435
chr9	84042684	84042761	CCATTCTCCCTGTCACTTTCAGGTACACCATTCAAACGTAGGTTTGGTCTTTTCTCATAGTCCCATATTTCTTGGAGG	1	436
chr9	95256012	95256089	CCTCCAAGAAATTCGGGACTATGTGAAAAGACAAAACCTACGTTTAATTGGTGTGTGGTGTACCTGAAAGTGACAAGG	1	437
chr9	101816988	101817065	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACCAGAAGAATGG	1	438
chr9	135842327	135842403	CCTCCAAGAAATATGGGACTATGTGAAAAGCCCAAACCTACGTTTGACTGATGTACCTAAAGTGACGGGGAGAATGG	3	439
chr9	136910865	136910940	CCCGCACTGTGAGCTTGGCCGAGTGCTGTCTGAAAGCATCCTTTCCCTTCACCTGGAGACTGGAGCGCCATAGAGG	4	440
chr10	13710312	13710389	CCTGTCTCCCCCATTCCATGCAAAATAAAACACAAACCAAGCTTTGCTTTAAGTGCTCCCTGATGCAGTTCAGCGTGG	1	441
chr10	18938129	18938206	CCATTCTTCCCGTCACATTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCCCATAGTCCCATATTTCTTAGAGG	1	442

chr10	22712838	22712914	CCCCCTGCTCAGCTTGGGGAAGAAAAATAC AAAAACGATGCTTTTAGGCATTTTAAACAAC TTCACTACATTGAGGG	2	443
chr10	22712838	22712913	CCCCCTGCTCAGCTTGGGGAAGAAAAATAC AAAAACGATGCTTTTAGGCATTTTAAACAAC TTCACTACATTGAGG	4	444
chr10	40160932	40161009	CCTTTGTGTTGTGTGTATTCAACTCACAGAG TGAAACCTTCCTTTATTCAGAGCAGTTTTGA AACACTCTTTTTGTGG	1	445
chr10	40390136	40390213	CCTTTGTGTTGTGTGTATTCAACTCACAGAG TGAAACCTTCCTTTATTCAGAGCAGTTTTGA AAAACACTTTTTGTGG	1	446
chr10	40409152	40409229	CCTTTGTGTTGTGTGTATTCAACTCACAGAG TGAAACCTTCCTTTATTCAGAGCAGTTTTGA AAAACTCTTTTTGTGG	1	447
chr10	40433940	40434017	CCTTTGTGTTGTGTGTATTCAACTCACAGAG TGAAACCTTCCTTTATTCAGAGCAGTTTTGA AACACTCTTTTTGTGG	1	448
chr10	40588155	40588232	CCTTTGTGTTGTGTGTATTCAACTCACAGAG TGAAACCTTCCTTTATTCAGAGCAGTTTTGA AATACTCTTTTTGTGG	1	449
chr10	41146207	41146284	CCTTTGTGTTGTGTGTATTCAACTCACAGAG TGAAACCTTCCTTTATTCAGAGCAGTTTTGA AACACTCTTTTTGTGG	1	450
chr10	43835183	43835260	CCATTCTCCCTGTCACTTTCAAGTACACCAA TCAAACCTAGGTTTGGTCTTTTCACATAGTTC CATATTTCTTGGAGG	1	451
chr10	54913222	54913299	CCCCTCCCATCACAGGCCCTGAGGTTTAAGA GAAAACCATGGTTTTGTGGGCCAGGCCCATG ACCCTTCTCCTCTGGG	1	452
chr10	54913222	54913298	CCCCTCCCATCACAGGCCCTGAGGTTTAAGA GAAAACCATGGTTTTGTGGGCCAGGCCCATG ACCCTTCTCCTCTGG	3	453
chr10	54913223	54913299	CCCTCCCATCACAGGCCCTGAGGTTTAAGAG AAAACCATGGTTTTGTGGGCCAGGCCCATGA CCCTTCTCCTCTGGG	2	454
chr10	54913223	54913298	CCCTCCCATCACAGGCCCTGAGGTTTAAGAG AAAACCATGGTTTTGTGGGCCAGGCCCATGA CCCTTCTCCTCTGG	4	455
chr10	58035951	58036028	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTCATCTTTTCACATAGTC CCACGGTTTTTGGAGG	1	456
chr10	58677525	58677602	CCTCCAAGATATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA ATTGATGGGGAGAATGG	1	457
chr10	84021390	84021467	CCTCCAAGAAATATGGGACTGTGTGAAAAG AACAAACCTACGTTTGATTGGTGTACGTGAA AGTGATGGGGAGAATGG	1	458
chr10	91442692	91442769	CCATTCCTCCCGTCACTTTCAGATACACCAA AAAAACGTAGGTTTGGTCTCTTCACATAGTC CCACATTTCTTGGAGG	1	459
chr10	91446848	91446925	CCTCCAAGAAATGTGGGACTATGTGAAGAG ACCAAACCTACGTTTTTTTGGTGTATCTGAA AGTGACGGGAGGAATGG	1	460
chr10	116928784	116928860	CCTCCAAGGGGAATCTGAGTTCTCTGAAGAC AAAAAGCATGGTTTCTTTTCTTCTGTATTTCT TATTGTTTCCTAGG	3	461
chr10	116937771	116937848	CCATTCTCCCTATCACTTTCCAGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	462
chr11	31182070	31182147	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTATACTTGAA ATTGACAAGGAGAATGG	1	463
chr11	34739273	34739350	CCTCCAAGAAATATGGGACTATGTGGAAAG ACCAAACCTACGTTTGACTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	464

chr11	86646529	86646606	CCTCTAAGAAATATGGGACTATGTGAAGAG ATGAAACCTACGTTTGATTGGTGTACCTGAA AGTGACGAGGAGAATGG	1	465
chr11	90469791	90469867	CCCTCGTATACTACATGCTATAGTCAAAGCA GTAAACCTTCCTTTCCTTAAGCAGACCACAC TCTTTCATGCCTGGG	3	466
chr11	90469792	90469867	CCTCGTATACTACATGCTATAGTCAAAGCAG TAAACCTTCCTTTCCTTAAGCAGACCACACT CTTTCATGCCTGGG	4	467
chr11	92429985	92430062	CCATTCTCCCCATCACTTTCAGGTATACTAAT CAAAGGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCATGGAGG	1	468
chr11	102818498	102818574	CCATTCCCCCGTCACTTTCAGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	2	469
chr11	120765065	120765142	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTTGTCTTTTCTTATAGTCC CATATTTCTTGGAGG	1	470
chr11	123131901	123131978	CCACTGCACCTGACCAAGATCCTTAATTTTT CTAAACCTACGTTTATCATCTATAAAATGAG CCATCTTTTCACATGG	1	471
chr11	129468520	129468597	CCTCCGAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGTTGTACCTGAA AGTGACAGGGAGAATGG	1	472
chr11	131272361	131272438	CCATTCTCCCCATCACTTTTAGGTACACCAA TCAAACGTAGGTTTGGTCCTTTTGCATAGAC CCATATTTCTTGGAGG	1	473
chr11	132761415	132761492	CCATTTTCCCCGTCAGTTTCATATACACCTAT CAAACGTAGGTTTACTGTTTTCACATAGTCC CTTATTTCTTGGAGG	1	474
chr12	22367416	22367493	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACCTTTGATTGGTGTACCTGAA AGTGACGGGCAGGATGG	1	475
chr12	33146384	33146461	CCATTCTTCTCGTCATTTTCAAGTACACCAAT CAAACGTAGGTTTGGTCTTTTCGCATAGTCC CATATTTCTTGGAGG	1	476
chr12	33198476	33198553	CCATTCTTCTCGTCACTTTCAAGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	477
chr12	46038332	46038409	CCTCCAAGAAATATAGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACTTGAA AGTGACAGGGAGAATGG	1	478
chr12	60236126	60236203	CCTCCAAGAAATGTGGAACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	479
chr12	62098359	62098434	CCCTGACACTGATAAACGGATATGAAGAGA AAAAAGCTAGGTTTTCGCTGGAATTCCTAAG CTTGGGCTGCAGTGG	4	480
chr12	62112591	62112668	CCCTTCTCCCAGTCACTTTTAGGTACACCAA TGAAACGTAGGTTTGGTCTTTTCACACAGTC CCATATTTCTTGGAGG	1	481
chr12	62112592	62112668	CCTTCTCCCAGTCACTTTTAGGTACACCAAT GAAACGTAGGTTTGGTCTTTTCACACAGTCC CATATTTCTTGGAGG	2	482
chr12	62418577	62418652	CCACTCCCTCTCCCCCAAAAAGTAAAGGTAG AAAACCAAGGTTTACAGGCAACAAATAGCA CAATGAATGGAATGG	4	483
chr12	71732311	71732388	CCAAACCCGCATCGCACACCCTGTGAGGGG GACAAAGGAACCTTTCCGTTCCAACATCAAG GTTGTTTTGACCCAAGG	1	484
chr12	78047816	78047893	CCATTCTTTCTGTCACTTTCAGGTATACCAGT CAAACCTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	485
chr12	81480016	81480093	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	486

chr12	96840231	96840307	CCACACGGTAGAGGATAAACTAGGTGGATT CTCAAAGCAACCTTTGAAATAATCTATGCAG TTTTTCTGGGTACTGG	3	487
chr12	99187165	99187242	CCACCAAGAAACATGGGACTATGTGAAAAG ACCAAACCTACGTTTGGTTGGTGTACCTGGA AGTGACGGGGAGAGTGG	1	488
chr12	107860841	107860918	CCTCCAAGAAATATGGGACCATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	489
chr12	110882809	110882885	CCTGTAAAAAGGTCACATGGTCAGGTGTGCC TAAACGATCCTTTTATTTATTTATTTATTTAT TTTTAAGAAACAGG	2	490
chr12	119063321	119063397	CCAGCCCCAAAATGTCAGGGGCTTAGAACA ACAAAGGTTCCTTTTCATGTTTATACTACAT GTTTGTCATGGGCTGG	2	491
chr13	35320704	35320781	CCGTTTTCCCCATCACTTTCAGGTACACCAG TCAAACGTAGGTTTGGTCTTTTCACATGGTC CCACATTTCTTGGAGG	1	492
chr13	53133477	53133554	CCTGGAATAGCTTTCCTGACTGTCTGACTTC AAAAACCTTGGTTTGACCACTTCGTCTATAT CATGAGGAAGGACTGG	1	493
chr13	53184880	53184956	CCCTACTCTGAACCTACCTTGATAAAGCCTA GAAAACCAAGCTTTGACAAGATTTGACAAG AGATGGAATTTGGAGG	3	494
chr13	53184881	53184956	CCTACTCTGAACCTACCTTGATAAAGCCTAG AAAACCAAGCTTTGACAAGATTTGACAAGA GATGGAATTTGGAGG	4	495
chr13	57896962	57897038	CCCTTATAAAACTGAAAACTTTAACCTTTTTT AAAGCATGCTTTTGAATAAATTCTTTTATTA CAAAAAAGACCAGG	2	496
chr13	62610100	62610177	CCATTCTCCCTGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACGTAGTC CCATATTTCTTGGAGG	1	497
chr13	77004382	77004458	CCCTTTATTATCCAAGTGGTTTCCTGCTCTTC AAACCTTCCTTTCAAAATTTTGTCTCCTACTT AAAACAAGTTAGG	2	498
chr13	81646075	81646151	CCTTCTGTTGAGACCTACTGCTAAGAAAACA AAAAAGGTTCCTTTCAAATATTATTGTGAAT CAATAATGTACCTGG	3	499
chr13	83755854	83755931	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTCATTGATGGACCTGAA AGTGATGGGGAGAATGG	1	500
chr13	89719199	89719275	CCATTCTCCCTTCACTTTCAGTTACACCAATC AAACGTAGGTTTGGTCTTTTCACATAGTCCC ATATTTCTTGGAGG	2	501
chr13	102010574	102010650	CCTAGGGAAGTGATCATAGCTGAGTTTCTGG AAAAACCTAGGTTTTAAAGTTGAGGAGACTT AAGTCCAAAACCTGG	3	502
chr13_KI270841v1_alt	124240	124316	CCATTCTCCCTTCACTTTCAGTTACACCAATC AAACGTAGGTTTGGTCTTTTCACATAGTCCC ATATTTCTTGGAGG	2	503
chr14	25980646	25980723	CCTCCAAGAAATATGGGACTATGTGAAAAG ACTAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	504
chr14	35842786	35842863	CCATTCTCCCTGTCACTTTCAGGTATGCCAGT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTCCTTGGAGG	1	505
chr14	42646400	42646477	CCTCCAAGAAATATGGGACTATGTAAAAAG ACGAAACCTACGTTTGATTGGTGTACTTAAA AGTGACGAGGAGAATGG	1	506
chr14	49063242	49063319	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTTGTGTACCTGAA AGTGATGGGGAGAATGG	1	507
chr14	49130379	49130456	CCATTCTCCCCGTCACTTTCAGGCACACCAA TCAAACGTAGGTTTAGTCTTTTCACATAGTC CCATATTTCTTAGAGG	1	508

chr14	51352342	51352418	CCTTAATGCATTCATATTTCATATTTTAAATA AAACCATGGTTTCCCACAGAGTGACTTCTAC TCTAAGAAATGGGG	2	509
chr14	51352342	51352417	CCTTAATGCATTCATATTTCATATTTTAAATA AAACCATGGTTTCCCACAGAGTGACTTCTAC TCTAAGAAATGGG	4	510
chr14	60835842	60835919	CCGTTCTTTCCGTCACTTTCAGGTACACCAGT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	511
chr14	66529072	66529148	CCATTCTCCCCATCACTTTCATGTACACCAAT CAAACGTAGGTTTGGTCTTTGTTAACATAGT CCCATATTTCTTGG	3	512
chr14	79210873	79210949	CCCTATAAAGCTTAGAGAAACACAGGGCTCT TTAAACGATCCTTTTTCTCTTTTCTGTTTTAA ATTTCATCACTTGG	3	513
chr14	79210874	79210949	CCTATAAAGCTTAGAGAAACACAGGGCTCTT TAAACGATCCTTTTTCTCTTTTCTGTTTTAAA TTTCATCACTTGG	4	514
chr14	85371541	85371618	CCATTCTCCCCATCACTTTCAGGTACACTAA TCAAAGGTAGGTTTGGTCTTTTCACATGGTC CTATATTTCTTGGAGG	1	515
chr14	92918713	92918790	CCCCATAGCACGATCACATGGGACATTCAGG GGAAAGCAACCTTTTCCAGGAAGGAAAACC CAATGCTGGGACCCAGG	1	516
chr14	92918714	92918790	CCCATAGCACGATCACATGGGACATTCAGG GGAAAGCAACCTTTTCCAGGAAGGAAAACC CAATGCTGGGACCCAGG	2	517
chr14	103386821	103386897	CCCTTTCAGCGCTCACAGGCTATGGTTTTAT AAAAGGAACCTTTGATTTTGTTCATGTGAAA CTACAAAATGCCAGG	2	518
chr14_KI270847v1_alt	33275	33352	CCCCATAGCACGATCACATGGGACATTCAGG GGAAAGCAACCTTTTCCAGGAAGGAAAACC CAATGCTGGGACCCAGG	1	519
chr14_KI270847v1_alt	33276	33352	CCCATAGCACGATCACATGGGACATTCAGG GGAAAGCAACCTTTTCCAGGAAGGAAAACC CAATGCTGGGACCCAGG	2	520
chr15	20630566	20630643	CCTCCAAGAAATATTGGAGTATGTGATAAGA CCAAACCTTCGTTTGACTGGTGTACCTGAAA GTGATGGGGAGAATGG	1	521
chr15	21675103	21675180	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	522
chr15	22117571	22117648	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	523
chr15	22369744	22369821	CCATTCTCCCCATCACTTTCAGGTACACCAG TCAAACGAAGGTTTGGTCTTATCACATACTC CAATATTTCTTGGAGG	1	524
chr15	42302832	42302909	CCTCCAAGATATATGGGACTATGTGAAAAG GCCAAACCTACCTTTGATTGATACACCTGAA AATGACAGGGAGAATGG	1	525
chr15	49967601	49967678	CCTCCAAGAAATATGCGACTATGTGAAAAG ACCAAACCTACGTTTCATTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	526
chr15	83964501	83964577	CCTCCAAGAAATATGGGACTATGTGGAAAG ACCAAACCTACGTTTGTTTGGTGTACCTGAA AGTGAGGGGAGAATGG	3	527
chr15	87261388	87261465	CCATTCTCCTCATCACTTTCAAGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC TTATATTTCTTGGAGG	1	528
chr15_KI270727v1_random	409348	409425	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	529
chr15_KI270851v1_alt	14235	14312	CCATTCTCCCCATCACTTTCAGGTACACCAG TCAAACGAAGGTTTGGTCTTATCACATACTC CAATATTTCTTGGAGG	1	530

chr15_KI270852v1_alt	440099	440176	CCATTCTCCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	531
chr16	22123671	22123748	CCAGCAGAAGAATCTGGGGCACAGTCTGTGAAAAAAGGTACCTTTCTTAAGCAGGGTTCTTATCCTTCATGGGTCTGG	1	532
chr16	25557623	25557700	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGTTGTACCTGAAAGTGAGGGGGAGAATGG	1	533
chr16	36427179	36427255	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	534
chr16	36476450	36476526	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	535
chr16	36512469	36512545	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	536
chr16	36520964	36521040	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACACAGCAGATTTGAAACACTGTTTTTCTGG	2	537
chr16	36524704	36524780	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	538
chr16	36566812	36566888	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	539
chr16	36573603	36573679	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	540
chr16	36667694	36667770	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	541
chr16	36677320	36677396	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	542
chr16	36683096	36683172	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	543
chr16	36691251	36691327	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	544
chr16	36710951	36711027	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	545
chr16	36750364	36750440	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	546
chr16	36791455	36791531	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACACAGCAGATTTGAAACACTGTTTTTCTGG	2	547
chr16	36856683	36856759	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	548
chr16	36926655	36926731	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	549
chr16	36931752	36931828	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	550
chr16	36948058	36948134	CCTTGTGTTGTGTGTATTCAACTCACCGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	551
chr16	36974541	36974617	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	552

chr16	36981331	36981407	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	553
chr16	36990839	36990915	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	554
chr16	37021075	37021151	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	555
chr16	37042812	37042888	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	556
chr16	37085971	37086047	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	557
chr16	37129462	37129538	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	558
chr16	37146110	37146186	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACACAGCAGATTTGAA ACACTGTTTTTCTGG	2	559
chr16	37157309	37157385	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	560
chr16	37183118	37183194	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	561
chr16	37190924	37191000	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	562
chr16	37221808	37221884	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	563
chr16	37259501	37259577	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	564
chr16	37272409	37272485	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	565
chr16	37281923	37281999	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGTA ACACTGTTTTTCTGG	2	566
chr16	37346472	37346548	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	567
chr16	37357000	37357076	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	568
chr16	37373301	37373377	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	569
chr16	37419498	37419574	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	570
chr16	37430714	37430790	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	571
chr16	37455845	37455921	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	572
chr16	37458558	37458634	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	573
chr16	37486127	37486203	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	574

chr16	37525183	37525259	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTGTGG	2	575
chr16	37536735	37536811	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	576
chr16	37554730	37554806	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	577
chr16	37575784	37575860	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	578
chr16	37577483	37577559	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	579
chr16	37583598	37583674	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	580
chr16	37696368	37696444	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTCCACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	581
chr16	37704524	37704600	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	582
chr16	37706223	37706299	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	583
chr16	37708941	37709017	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	584
chr16	37763622	37763698	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	585
chr16	37772115	37772191	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	586
chr16	37791815	37791891	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	587
chr16	37796229	37796305	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	588
chr16	37797928	37798004	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	589
chr16	37843453	37843529	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	590
chr16	37848548	37848624	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	591
chr16	37864846	37864922	CCTTGTGTTGTGTGTATTCAACTCACCGAGTT AAACGATCCTTTACACAGAGCAGATTTGAAA CACTGTTTTTCTGG	2	592
chr16	37902550	37902626	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	593
chr16	37907307	37907383	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	594
chr16	37928033	37928109	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	595
chr16	37959262	37959338	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	596

chr16	37964355	37964431	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	597
chr16	37974881	37974957	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA AAACTGTTTTTCTGG	2	598
chr16	37987789	37987865	CCTTGTGTTGTGTGTATTCAACTCACAGAGT AAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	599
chr16	37994586	37994662	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTGTGG	2	600
chr16	38006479	38006555	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	601
chr16	38011567	38011643	CCTTGTGTTGTGTGTATTTAACTCACAGAGTT AAACGATCCTTTACACAGAGCAGATTTGAAA CACTGTTTTTCTGG	2	602
chr16	38040096	38040172	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	603
chr16	38041456	38041532	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	604
chr16	38062179	38062255	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	605
chr16	38102937	38103013	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	606
chr16	38128412	38128488	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	607
chr16	38131809	38131885	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	608
chr16	38144723	38144799	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	609
chr16	38168845	38168921	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	610
chr16	38209287	38209363	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	611
chr16	38210986	38211062	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	612
chr16	38229667	38229743	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGATTTGAA ACACTGTTTTTCTGG	2	613
chr16	47424037	47424114	CCATTCTCCCTATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	614
chr16	60730549	60730625	CCTCGTCACTGCCAGATTTTGTGGCTACCAG CAAAGGATCGTTTTAAGCTGCAACTCAGGAA ATTGAGAAAATATGG	2	615
chr16	72545014	72545091	CCTCCAAGAAATATGGGACTATGTGAAAAA ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	616
chr16	81945503	81945579	CCCTGTGTTCTTTTATACTAAAACAAGCCAG CAAACCAACCTTTGAGATGTGTTGCCTTAAA CATTACTGAATGGGG	2	617
chr16	81945503	81945578	CCCTGTGTTCTTTTATACTAAAACAAGCCAG CAAACCAACCTTTGAGATGTGTTGCCTTAAA CATTACTGAATGGG	4	618

chr17	16474024	16474100	CCGAGAAACGGCTTTAGCAACAAATAAATA TCAAAAGGATGCTTTCTCTTCAGAATAATCT AAAGTAAGTTGGGAGG	3	619
chr17	34438512	34438589	CCATGTTACTCCGGATAAGGACAGCAAAGG AGGAAAGGAACCTTTTCTGGGCCACCAGAA GGATGAGCTTGGGCTTGG	1	620
chr17	43690782	43690859	CCCAGGGATATGCTGGCCACGGGGAGGAGC CGGAAACCAACCTTTGTGTCACTGTGTAGTG ACAAGTGCCTTTGGAGG	1	621
chr17	43690783	43690859	CCAGGGATATGCTGGCCACGGGGAGGAGCC GGAAACCAACCTTTGTGTCACTGTGTAGTGA CAAGTGCCTTTGGAGG	2	622
chr17	69156298	69156375	CCTTAGGGACCCATAATGGCCACAACCAGG AGAAAAGCAAGCTTTGATGCTTAAACACTAC TTACAGACATGTACAGG	1	623
chr17	74595228	74595305	CCTGCCTCTGTTCCTCCTTCCTGATGGTGGCG GAAAGGATGCTTTTGCCAGATCAACAGTCAC ACACAACACACCAGG	1	624
chr17	83191644	83191721	CCTGACTCCAGCCCTCCTTGACAAGGTCTCC GTAAAGCATGCTTTCTCTTAGGGACCCTCAG AGGGAGGCTTGGTGGG	1	625
chr17	83191644	83191720	CCTGACTCCAGCCCTCCTTGACAAGGTCTCC GTAAAGCATGCTTTCTCTTAGGGACCCTCAG AGGGAGGCTTGGTGG	3	626
chr18	35135224	35135300	CCTTATTTGGAATGTGACAAGACCCATTTGT TTAAACCTTGGTTTTTATGCAGAAAGAAAAG GAAGGCTGCAGTGGG	3	627
chr18	38918861	38918938	CCATTCTCCCTGTCACTTTCAGGTACACTAAT CAAACGTAGGTTTGCTGTTTTTACATAGGCT CATATTTCTTGGAGG	1	628
chr18	45476589	45476666	CCATTCTCCCCATCACTTTCAGGTACACCAG TCAAACGTAGGTTTGGTCTTTTCACATAGTC CCATATTTCTTGGAGG	1	629
chr18	48640821	48640896	CCTGTTTGTTATTTTAGCTAATGTCAAAAAG AAAACCTTGCTTTTTCTGAACCCTTTCAGAG GCAGAAAGTGGGGG	4	630
chr18	71096732	71096808	CCATTTTCCCCACCACTTTCACGTACAGCAA TCAAACGTAGGTTTGGTCTTTTCACTAGTCC CATATTTCTTGGAGG	3	631
chr19	24957844	24957920	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTGTTGTGG	2	632
chr19	25015316	25015392	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCATACTTGAA ACACTCTTTTTGTGG	2	633
chr19	25074119	25074195	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	634
chr19	25827861	25827937	CCTTGTGTTGTGTTTATTCAACTCACAGAGTT AAACGATCCTTTACACAGAGCAGACTTGAA ATACTCTTTTTGTGG	2	635
chr19	26054056	26054132	CCTTGTAGTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCATACTTGAA ACACTCTTTTTGTGG	2	636
chr19	26211777	26211853	CCTTGTATTGTGAGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	637
chr19	26483670	26483746	CCTTGTGTTGTGTGTCTTCAACTCACAGAGTT AAACGATGCTTTACACAGAGTAGACTTGAA ACACTCTTTTTCTGG	2	638
chr19	26636516	26636592	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGTA ACACTCTTTTTGTGG	2	639
chr19	26637877	26637953	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACGTGA AACACTCTTTTTGTGG	2	640

chr19	26750223	26750299	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGCAGACTTGAA ACACTCTTTTTGTGG	2	641
chr19	26841158	26841234	CCTTGTGTTGTGTGTATTCAACTCACAGAGT TAAACGATCCTTTACACAGAGGAGACTTGTA ACACTCTTTTTGTGG	2	642
chr19	28517220	28517297	CCAGGAAAAAATTTAAACTTTCTTAACTTGA TAAAAGGTAGCTTTCAAAACCTACAATAAAT AACATACTTAGAGTGG	1	643
chr19	34566821	34566898	CCATTCTCCTCGTCACTTTCAGGTACACCAA ACAAACGTAGGTTTGGTCTTTTTACGTAGTC CCATATTTCTTGGAGG	1	644
chr19	52261770	52261847	CCCTCTTGAAGTTAGGGAAGTAGCATTTAAG GGAAACGTAGCTTTACTATTAAGAATTTCAA ACAGCACTTGTCAGGG	1	645
chr19	52261770	52261846	CCCTCTTGAAGTTAGGGAAGTAGCATTTAAG GGAAACGTAGCTTTACTATTAAGAATTTCAA ACAGCACTTGTCAGG	3	646
chr19	52261771	52261847	CCTCTTGAAGTTAGGGAAGTAGCATTTAAGG GAAACGTAGCTTTACTATTAAGAATTTCAAA CAGCACTTGTCAGGG	2	647
chr19	52261771	52261846	CCTCTTGAAGTTAGGGAAGTAGCATTTAAGG GAAACGTAGCTTTACTATTAAGAATTTCAAA CAGCACTTGTCAGG	4	648
chr20	11151392	11151469	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCACATATTCC CATATTTCTTGGAGG	1	649
chr20	14027067	14027143	CCATTCTCCCTTCACTTTCAGGTACACCAATC AAACGTAGGTTTGGTCTTTTCACATAGTCCC ATATTTTTTGGAGG	2	650
chr20	50615399	50615476	CCTATAGTCTCAGTTACTTGGGAGGCTGAGG TAAAAGGATCGTTTGAGCCCAGGAGGTGGA GGTTGCAGTGAGCCGGG	1	651
chr20	50615399	50615475	CCTATAGTCTCAGTTACTTGGGAGGCTGAGG TAAAAGGATCGTTTGAGCCCAGGAGGTGGA GGTTGCAGTGAGCCGG	3	652
chr20	60909414	60909490	CCTTTCCCAACTCTGCTATTGCCCCCACATCC TAAAGGAACCTTTCTTTTTTTATATATTTTAT TTTAAGTTCCAGG	3	653
chr21	16226086	16226163	CCTCCAAGAAATATGGAACTATGTGAAAAG ACCAAACCTACGTTTGATTGACGTACCTGAA AGTGACAGGGAGAATGG	1	654
chr21	17835234	17835309	CCTCTTCTGAAAGCATTGATAATCAACATTT TAAACGTAGCTTTTCCCCATATTGCTAGGAA GGCTCATTCCCGGG	4	655
chr21	19425636	19425713	CCTCCAAGAAATATGGGACTATGTGAAAAG GCCAAACCTACGTTTGATTGCTGTACCCGAG AGTGACGGGGAGAATGG	1	656
chr21	32220958	32221035	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	657
chr21	34335877	34335953	CCCGGGGCCTGGGTGCCCAGTGCCAGTGGTC AGAAAGGTTGCTTTGGTGTTTTTCATTGTTA GTGAGACAGAGATGG	3	658
chr21	34335878	34335953	CCGGGGCCTGGGTGCCCAGTGCCAGTGGTCA GAAAGGTTGCTTTGGTGTTTTTCATTGTTAGT GAGACAGAGATGG	4	659
chr21	36315276	36315353	CCATTCTCCCCATCATTTTCAGGTACACCAA TCAAACGTAGGTTTGATCTTTTCACATAGCC CCATATTTCTTGGAGG	1	660
chr21	41547952	41548028	CCACCAGCACTTCTGTTAGAAGTTGCAGCAG AGAAAGGATCCTTTAGGCACATCTCCCAGAT CCTTGCGAAGAGGGG	3	661
chr22	18973194	18973271	CCTGTGCCAGGGTCCTTCCACTGGGACTGGC AGAAACGTAGGTTTGCATGGAGTGAGAAGC AGGGGAGAGGTTGAGGG	1	662

chr22	18973194	18973270	CCTGTGCCAGGGTCCTTCCACTGGGACTGGC AGAAACGTAGGTTTGCATGGAGTGAGAAGC AGGGGAGAGGTTGAGG	3	663
chr22	20265462	20265539	CCCTCAGCCTCTCCCCTGCTTCTCACTCCATG CAAACCTACGTTTCTGCCAGTCCCAGCAGAA GGACCCTGGCACGGG	1	664
chr22	20265462	20265538	CCCTCAGCCTCTCCCCTGCTTCTCACTCCATG CAAACCTACGTTTCTGCCAGTCCCAGCAGAA GGACCCTGGCACGG	3	665
chr22	20265463	20265539	CCTCAGCCTCTCCCCTGCTTCTCACTCCATGC AAACCTACGTTTCTGCCAGTCCCAGCAGAAG GACCCTGGCACGGG	2	666
chr22	20265463	20265538	CCTCAGCCTCTCCCCTGCTTCTCACTCCATGC AAACCTACGTTTCTGCCAGTCCCAGCAGAAG GACCCTGGCACGG	4	667
chrX	27300998	27301075	CCTCCAAGAAATATGGGGCTATGTGAAAAG ACCAAACCTACCTTTGATTGGTGTATCTGAA AGTGACGGGGAGAATGG	1	668
chrX	28456666	28456743	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTTGTGTACCTGAA AGTGATGGGGAGAATGG	1	669
chrX	35634985	35635062	CCATTCTCCCCGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTCTTTTCTCATTGTCC CATATTTCTTGGAGG	1	670
chrX	39460148	39460223	CCCATCAAGAGCGGTTGTGCATGGCAACAGT AAAAGGATGGTTTGTTACACTAGTACAAAA AGAGGTGGCCAGAGG	4	671
chrX	43926403	43926480	CCATTCTCTCTGTCACTTTCAGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	672
chrX	44254600	44254677	CCTCCAAGAAATACGGGACTATGTGAAAAG ACCAAACGTACGTTTGATTGGTGTACCTGAA AGTGATAGGGAGAATGG	1	673
chrX	46088602	46088679	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACTGGGAGAATGG	1	674
chrX	50222874	50222951	CCATTCTCCCTGTCACTTTCAGGTACACGAA TCAAACGTAGGTTTCATCTTTTCACATAGTC CCATATTTCTTAGAGG	1	675
chrX	57416835	57416911	CCATTCTCTCTGTCACTTTCTGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTTT CACATATTTCTTGG	3	676
chrX	57856466	57856543	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAAGGAAAATGG	1	677
chrX	62702479	62702556	CCTGAAAAACATTGTTTCCAACCTGGTAAAT CAAAAGGAAGGTTTAACTTTGTTAGATAAGT CCACATATCACCAAGG	1	678
chrX	63067129	63067206	CCTCCAAGAAATGTGGGACTATGGGAAAAG ACCAAACCTACCTTTGTTTGGTGTACCTGAA AGTGACGGGGAGAAAGG	1	679
chrX	64936250	64936327	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTCATTGGTGTACCTGAA AGTGATGGGTAGAATGG	1	680
chrX	66720099	66720176	CCTACAAGAAATATGGGACTATGGGAAAAG ACCAAACCTACGTTTGATTGGTACACTGGAA AGTGACAGGGATAATGG	1	681
chrX	68529086	68529163	CCATTCTCCCTGTCACTTTCTGGTACACCAAT CAAAGGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	682
chrX	73893994	73894071	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGTGTACCTGAA AGTGATGGGGAGAATGG	1	683
chrX	75723201	75723278	CCATTCTCTTTGTCACTTTCAGGTATACCAAT CAAACGTTGGTTTGGTCTTTTTGCATAGTCCC ATATTTTGTGGAGG	1	684

chrX	75815659	75815736	CCTCCAAGAAATATGAGACTATGTGAAAAG ACCAAACCTACGTTTGATTAGTGTACCTGAA AATGATGGGGAGAATGG	1	685
chrX	80967103	80967180	CCATTCTTTCTGTCACTTTCAGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	686
chrX	89936425	89936502	CCATTCTCCCTGTCACTTTCAGGTACACCAA TCAAACGTAGGTTTGTTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	687
chrX	91038768	91038845	CCATTATCCCCATCACTTTCAGGTACACCAA TCAAACGTAGGTTTGGTTTTTTCACATAGTTC AATATTTCTTTGAGG	1	688
chrX	91471271	91471348	CCTCCAAGAAATATGGGACTATCTGAAAAG ATCAAACCTACGTTTGATTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	689
chrX	96428180	96428257	CCTTTCTCCCCATCACTTTCAGGTACACCAAT CAAACGTAGGTTTGGTCTTTTCATATAGTCC CATATTTCTTGGAGG	1	690
chrX	100268291	100268368	CCTCCAAGAAATATGGGACTATGTGCAAAG ATCAAACCTACGTTTGATTGCTGTACCTGAA AGTGATGGGGAGAATGG	1	691
chrX	105811046	105811123	CCATTCTCCCCATCACTTTCAGGTACACCAG TCAAACGTAGGTTTGGTCTTTTCACATAATC CCATATTTCTTGGAGG	1	692
chrX	115673065	115673141	CCTCCAAGAAGTATGGGACCATGGAAAAGA TCAAACCTACGTTTGACTGGTGTACCTGAAA GTGACTGGGAGAATGG	2	693
chrX	117269846	117269923	CCTCCAAGAAATATGGGACTATGTGAAAAG ACCAAACCTACGTTTGATTGGAGTACTTGAA AATGACAGGGATAATGG	1	694
chrX	139191369	139191445	CCTTTAAAGACATGCTCTTTGTGCCAGAAAT TCAAAGGTTGCTTTTATGTCCAGTGGGGTGG AGGGAGGAAGCTCGG	3	695
chrX	147988614	147988691	CCATTCTCCCCGTCACTTTCAGGGACCTCAA TCAAACGTAGGTTTTGTCTTTTCACATAGTCC CATATTTCTTGGAGG	1	696
chrX	155321041	155321118	CCTCCAAGAAATATAGGACTATGTGAAAAG ACCAAACCTACGTTTGACTGGTGTACCTGAA AGTGACAGGGAGAATGG	1	697
chrY	15109391	15109468	CCATTCTCCCCATCACTTTCAGGTACACCAA TCAAAGGTAGGTTTGGTCTTTTCACATAGTC CGATATTTCCTGCAGG	1	698

Chromosomal sites were identified by searching for $CCX_{(30-31)}$-AAASSWWSSTTT-$X_{(30-31)}$-GG (SEQ ID NO: 699) where W is T or A and S is G or C. Pattern 1 is $CCX_{(31)}$-AAASSWWSSTTT-$X_{(31)}$-GG (SEQ ID NO: 699), 2 is $CCX_{(30)}$-AAASSWWSSTTT-$X_{(31)}$-GG (SEQ ID NO: 699), 3 is $CCX_{(31)}$-AAASSWWSSTTT-$X_{(30)}$-GG (SEQ ID NO: 699), and 4 is $CCX_{(30)}$-AAASSWWSSTTT-$X_{(30)}$-GG (SEQ ID NO: 699). Only the + strand is shown and the start and end corresponds to the first and last base pair in the chromosome (GRCh38) or alternate assembly when applicable.

DNA sequencing

[00165] Transfections of 293T cells were performed as above in sextuplet and incubated for 72 hours. Cells were harvested and replicates were combined. Episomal DNA was extracted using a modified HIRT extraction involving alkaline lysis and spin column purification essentially as described (Quan *et al.*, Circular polymerase extension cloning of complex gene libraries and pathways. PloS one 4, e6441 (2009); and Hillson (2010), vol. 2015, pp. CPEC protocol; the entire contents of each of which are hereby incorporated by reference). Briefly, after harvesting, HEK293T cells were washed in 500 µL of ice cold PBS,

resuspended in 250 µL GTE Buffer (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA and pH 8.0), incubated at room temperature for 5 minutes, and lysed on ice for 5 minutes with 200 µL lysis buffer (200 mM NaOH, 1% sodium dodecyl sulfate). Lysis was neutralized with 150 µL of a potassium acetate solution (5 M acetate, 3 M potassium, pH 6.7). Cell debris were pelleted by centrifugation at 21,130 *g* for 15 minutes and lysate was applied to Econospin Spin columns (Epoch Life Science, Missouri City, TX). Columns were washed twice with 750 µL wash buffer (Omega Bio-tek, Norcross, GA) and eluted in 45 µL TE buffer, pH 8.0.

[00166] Isolated episomal DNA was digested for 2 hours at 37 °C with RecBCD (10 U) following the manufacturer's instructions and purified into 10 µL EB with a MinElute Reaction Cleanup Kit (Qiagen, Valencia, CA). Mach1-T1 chemically competent cells were transformed with 5 µL of episomal extractions and plated on agarose plates selecting for carbenicillin resistance (containing 50 µg/mL carbenicillin). Individual colonies were sequenced with primer pCALNL-for-1 to determine the rate of recombination. Sequencing reads revealed either the 'left' intact non-recombined recCas9 site, the expected recombined product, rare instances of 'left' non-recombined site with small indels, or one instance of a large deletion product.

Analysis of recCas9 catalyzed genomic deletions

[00167] HEK293T cells were seeded at a density of $6x10^5$ cells per well in 24 well collagen-treated plates and grown overnight (Corning, Corning, NY). Transfections reactions were brought to a final volume of 100 µL in Opti-MEM (ThermoFisher Scientific, Waltham, MA). For each transfection, 90 ng of each guide RNA expression vector, 20 ng of pmaxGFP (Lonza, Allendale, NJ) and 320 ng of recCas9 expression vector were combined with 2 µL Lipofectamine 2000 in Opti-MEM (ThermoFisher Scientific, Waltham, MA) and added to individual wells. After 48 hours, cells were harvested and sorted for the GFP transfection control on a BD FACS AriaIIIu cell sorter. Cells were sorted on purity mode using a 100 µm nozzle and background fluorescence was determined by comparison with untransfected cells. Sorted cells were collected on ice in PBS, pelleted and washed twice with cold PBS. Genomic DNA was harvested using the E.Z.N.A. Tissue DNA Kit (Omega Bio-Tek, Norcross, GA) and eluted in 100 µL EB. Genomic DNA was quantified using the Quant-iT PicoGreen dsDNA kit (ThermoFisher Scientific, Waltham, MA) measured on a Tecan Infinite M1000 Pro fluorescence plate reader.

[00168] Nested PCR was carried out using Q5 Hot-Start Polymerase 2x Master Mix supplemented with 3% DMSO and diluted with HyClone water, molecular biology grade (GE Life Sciences, Logan, UT). Primary PCRs were carried out at 25 uL scale with 20 ng of genomic DNA as template using the primer pair FAM19A2-F1 and FAM19A2-R1 (Table 5). The primary PCR conditions were as follows: 98°C for 1 minute, 35 cycles of (98°C for 10 seconds, 59°C for 30 seconds, 72°C for 30 seconds), 72°C for 1 minute. A 1:50 dilution of the primary PCR served as template for the secondary PCR, using primers FAM19A2-F2 and FAM19A2-R2. The secondary PCR conditions were as follows: 98°C for 1 minute, 30 cycles of (98°C for 10 seconds, 59°C for 20 seconds, 72°C for 20 seconds), 72°C for 1 minute. DNA was analyzed by electrophoresis on a 1% agarose gel in TAE alongside a 1 Kb Plus DNA ladder (ThermoFisher Scientific, Waltham, MA). Material to be Sanger sequenced was purified on a Qiagen Minelute column (Valencia, CA) using the manufacturer's protocol. Template DNA from 3 biological replicates was used for three independent genomic nested PCRs.

[00169] The limit of detection was calculated given that one complete set of human chromosomes weighs approximately 3.6 pg ($3.3 \cdot 10^9$ bp $\times$ $1 \cdot 10^{-21} \frac{g}{bp}$). Therefore, a PCR reaction seeded with 20 ng of genomic DNA template contains approximately 5500 sets of chromosomes.

[00170] For quantification of genomic deletion, nested PCR was carried out using the above conditions in triplicate for each of the 3 biological replicates. A two-fold dilution series of genomic DNA was used as template, beginning with the undiluted stock (for sample 1, 47.17 ng/uL; for sample 2, 75.96 ng/uL; and for sample 3, 22.83 ng/uL) to reduce potential sources of pipetting error. The lowest DNA concentration for which a deletion PCR product could be observed was assumed to contain a single deletion product per total genomic DNA.

[00171] The number of genomes present in a given amount of template DNA can be inferred, and thus an estimate a minimum deletion efficiency for recCas9 at the *FAM19A2* locus can be determined. For example, take the case of a two-fold dilution series, beginning with 20 ng genomic DNA template. After nested PCR, only the well seeded with 20 ng yielded the correct PCR product. At 3.6 pg per genome, that PCR contained approximately 5500 genomes, and since at least one recombined genome must have been present, the minimum deletion efficiency is 1 in 5500 or 0.018%.

[00172] The levels of genomic DNA were quantified using a limiting dilution of genomic template because using quantitative PCR (qPCR) to determine the absolute level of

genome editing would require a set of PCR conditions that unambiguously and specifically amplify only from post-recombined genomic DNA. As shown in Figure 5B, primary PCR using genomic DNA as a template results in a roughly 2.5 kb off-target band as the dominant species; a second round of PCR using nested primers is required to reveal guide RNA- and recCas9-dependent genome editing.

Results

Fusing Gin recombinase to dCas9

[00173] It has been recently demonstrated that the N-terminus of dCas9 may be fused to the FokI nuclease catalytic domain, resulting in a dimeric dCas9-FokI fusion that cleaved DNA sites flanked by two guide RNA-specified sequences (see, *e.g.*, Guilinger *et al.*, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature biotechnology, (2014); Tsai *et al.*, Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature biotechnology, (2014); the entire contents of each of which are hereby incorporated by reference). The same fusion orientation was used to connect dCas9 to Ginβ, a highly active catalytic domain of dimeric Gin invertase previously evolved by Barbas and co-workers (Gaj *et al.*, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic acids research 41, 3937-3946 (2013), the entire contents of which is hereby incorporated by reference). Ginβ promiscuously recombines several 20-bp core "gix" sequences related to the native core sequence CTGTAAACCGAGGTTTTGGA (SEQ ID NO: 700) (Gaj *et al.*, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic acids research 41, 3937-3946 (2013); Klippel *et al.*, The DNA Invertase Gin of Phage Mu - Formation of a Covalent Complex with DNA Via a Phosphoserine at Amino-Acid Position-9. Embo Journal 7, 1229-1237 (1988); Mertens *et al.*, Site-specific recombination in bacteriophage Mu: characterization of binding sites for the DNA invertase Gin. The EMBO journal 7, 1219-1227 (1988); Plasterk *et al.*, DNA inversions in the chromosome of Escherichia coli and in bacteriophage Mu: relationship to other site-specific recombination systems. Proceedings of the National Academy of Sciences of the United States of America 80, 5355-5358 (1983); the entire contents of each of which are hereby incorporated by reference). The guide RNAs localize a recCas9 dimer to a gix site flanked by two guide-RNA specified sequences, enabling the Ginβ domain to catalyze DNA recombination in a guide RNA-programmed manner (Figure 1D).

[00174] To assay the resulting dCas9-Ginβ (recCas9) fusions, a reporter plasmid containing two recCas9 target sites flanking a poly-A terminator that blocks *EGFP* transcription was constructed (Figures 1A-1C). Each recCas9 target site consisted of a gix core pseudo-site flanked by sites matching a guide RNA protospacer sequence. Recombinase-mediated deletion removed the terminator, restoring transcription of *EGFP*. HEK293T cells were cotransfected with this reporter plasmid, a plasmid transcribing a guide RNA(s), and a plasmid producing candidate dCas9-Ginβ fusion proteins, and the fraction of cells exhibiting EGFP fluorescence was used to assess the relative activity of each fusion construct.

[00175] Parameters influencing the architecture of the recCas9 components, including the spacing between the core gix site and the guide RNA-binding site (from 0 to 7 bp), as well as linker length between the dCas9 and Ginβ moieties ($(GGS)_2$ (SEQ ID NO: 182), $(GGS)_5$ (SEQ ID NO: 701), or $(GGS)_8$ (SEQ ID NO: 183)) were varied (Figures 2A-2F). Most fusion architectures resulted in no observable guide RNA-dependent *EGFP* expression (Figures 1C-1D). However, one fusion construct containing a linker of eight GGS repeats and 3- to 6-base pair spacers resulted in approximately 1% recombination when a matched, but not mismatched, guide RNA was present (Figures 2E-2F). Recombination activity was consistently higher when 5-6 base pairs separated the dCas9 binding sites from the core (Figure 2F). These results collectively reveal that specific fusion architectures between dCas9 and Ginβ can result in guide RNA-dependent recombination activity at spacer-flanked gix-related core sites in human cells. The 8xGGS linker fusion construct is referred to as "recCas9".

Targeting DNA sequences found in the human genome with recCas9

[00176] Low levels of observed activity may be caused by a suboptimal guide RNA sequence or core gix sequence, consistent with previous reports showing that the efficiency of guide RNA:Cas9 binding is sequence-dependent (see, *e.g.*, Xu *et al.*, Sequence determinants of improved CRISPR sgRNA design. Genome research 25, 1147-1157 (2015), the entire contents of which is hereby incorporated by reference). Moreover, although the present optimization was conducted with the native gix core sequence (see, *e.g.*, Klippel *et al.*, The DNA Invertase Gin of Phage Mu - Formation of a Covalent Complex with DNA Via a Phosphoserine at Amino-Acid Position-9. Embo Journal 7, 1229-1237 (1988); Mertens *et al.*, Site-specific recombination in bacteriophage Mu: characterization of binding sites for the DNA invertase Gin. The EMBO journal 7, 1219-1227 (1988); Plasterk *et al.*, DNA

inversions in the chromosome of Escherichia coli and in bacteriophage Mu: relationship to other site-specific recombination systems. Proceedings of the National Academy of Sciences of the United States of America 80, 5355-5358 (1983); the entire contents of each of which are hereby incorporated by reference), several studies have shown that zinc finger-Gin or TALE-Gin fusions are active, and in some cases more active, on slightly altered core sites. See, *e.g.*, Gordley *et al.*, 3rd, Synthesis of programmable integrases. Proceedings of the National Academy of Sciences of the United States of America 106, 5053-5058 (2009); Gersbach *et al.*, Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. Nucleic acids research 39, 7868-7878 (2011); Mercer *et al.*, Chimeric TALE recombinases with programmable DNA sequence specificity. Nucleic acids research 40, 11163-11172 (2012); Gaj *et al.*, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic acids research 41, 3937-3946 (2013); Gordley *et al.*, 3rd, Evolution of programmable zinc finger-recombinases with activity in human cells. J Mol Biol 367, 802-813 (2007); Gersbach *et al.*, 3rd, Directed evolution of recombinase specificity by split gene reassembly. Nucleic acids research 38, 4198-4206 (2010); and Gaj *et al.*, Structure-guided reprogramming of serine recombinase DNA sequence specificity. Proceedings of the National Academy of Sciences of the United States of America 108, 498-503 (2011); the entire contents of each of which are hereby incorporated by reference). Thus, sequences found within the human genome were targeted in order to test if unmodified human genomic sequences were capable of being targeted by recCas9 and to test if varying the guide RNA and core sequences would increase recCas9 activity.

[00177] To identify potential target sites, previous findings that characterized evolved Gin variants (see, *e.g.*, Gaj *et al.*, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic acids research 41, 3937-3946 (2013), the entire contents of which is hereby incorporated by reference) as well as the observations above were used. Using this information, the human genome was searched for sites that contained $CCN_{(30-31)}$-AAASSWWSSTTT-$N_{(30-31)}$-GG (SEQ ID NO: 699), where W is A or T, S is G or C, and N is any nucleotide. The $N_{(30-31)}$ includes the N of the NGG protospacer adjacent motif (PAM), the 20-base pair Cas9 binding site, a 5- to 6-base pair spacing between the Cas9 and gix sites, and the four outermost base pairs of the gix core site. The internal 12 base pairs of the gix core site (AAASSWWSSTTT, SEQ ID NO: 699) were previously determined to be important for Ginβ activity(see, *e.g.*, Gaj *et al.*, Nucleic acids research 41, 3937-3946 (2013).

[00178] The search revealed approximately 450 such loci in the human genome (Table 9). A reporter construct was created, containing the sequence identical to one of these genomic loci, found in *PCDH15*, and then guide RNA expression vectors were constructed to direct recCas9 to this sequence (Figure 3A). These vectors encoded two pairs of guide RNAs, each of which contain spacer sequences that match the 5' and 3' regions flanking the *PCDH15* psuedo gix sites. Co-transfection of the reporter plasmid, combinations of these flanking guide RNA expression vectors, and the recCas9 expression vector resulted in *EGFP* expression in 11%-13% of transfected cells (Figure 3B), representing a > 10-fold improvement in activity over the results shown in Figure 2. These findings demonstrate that a more judicious choice of recCas9 target sequences can result in substantially improved recombination efficiency at DNA sequences matching those found in the human genome.

[00179] Next, whether both guide RNA sequences were required to cause recCas9-mediated deletion was determined. HEK293T cells were co-transfected with just one of the guide RNA vectors targeting the 5' or 3' flanking sequences of the *PCDH15* psuedo-gix core site, the *PCDH15* reporter plasmid, and a recCas9 expression vector. These co-transfections resulted in 2.5-3% EGFP expression (Figure 3B). The low levels of activity observed upon expression of just one of the targeting guide RNAs and recCas9 may be caused by the propensity of hyperactivated gix monomers to form dimers (see, *e.g.*, Gaj *et al.*, Enhancing the Specificity of Recombinase-Mediated Genome Engineering through Dimer Interface Redesign. J Am Chem Soc 136, 5047-5056 (2014), the entire contents of which is hereby incorporated by reference); transient dimerization may occasionally allow a single protospacer sequence to localize the dimer to a target site. No activity was detected above background when using off-target guide RNA vectors or when the recCas9 vector was replaced by pUC (Figure 3B).

[00180] These findings demonstrate that recCas9 activity can be increased substantially over the modest activity observed in the initial experiments by choosing different target sites and matching guide RNA sequences. A greater than 10-fold increase in activity on the *PCDH15* site compared to the original target sequences was observed (compare Figure 3B with Figure 2F). Further, maximal recombination activity is dependent on the presence of both guide RNAs and recCas9.

Orthogonality of recCas9

[00181] Next, whether recCas9 could target multiple, separate loci matching sequences found in the human genome in an orthogonal manner was tested. A subset of the recCas9

target sites in the human genome based on their potential use as a safe-harbor loci for genomic integration, or in one case, based on their location within a gene implicated in genetic disease, were selected.

[00182] To identify these sites, ENSEMBL (release 81) was searched to identify which predicted recCas9 target sites fall within annotated genes (see, *e.g.*, Cunningham *et al.*, Ensembl 2015. Nucleic acids research 43, D662-669 (2015), the entire contents of which is hereby incorporated by reference). One such site fell within an intronic region of *FGF14*. Mutations within *FGF14* are believed to cause spinocerebellar ataxia 27 (SCA 27) (see, *e.g.*, van Swieten *et al.*, A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected]. Am J Hum Genet 72, 191-199 (2003); Brusse *et al.*, Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14 gene (SCA27): A new phenotype. Mov Disord 21, 396-401 (2006); Choquet *et al.*, A novel frameshift mutation in FGF14 causes an autosomal dominant episodic ataxia. Neurogenetics 16, 233-236 (2015); Coebergh *et al.*, A new variable phenotype in spinocerebellar ataxia 27 (SCA 27) caused by a deletion in the FGF14 gene. Eur J Paediatr Neurol 18, 413-415 (2014); Shimojima *et al.*, Spinocerebellar ataxias type 27 derived from a disruption of the fibroblast growth factor 14 gene with mimicking phenotype of paroxysmal non-kinesigenic dyskinesia. Brain Dev 34, 230-233 (2012); the entire contents of each of which are incorporated herein by reference). Finally, a fraction of the predicted recCas9 target sites that did not fall within genes were manually interrogated to determine if some sequences fell within safe harbor loci. Using annotations in ENSEMBL genomic targets that matched most of the five criteria for safe harbor loci described by Bushman and coworkers were identified (Cunningham *et al.*, Ensembl 2015. Nucleic acids research 43, D662-669 (2015); and Sadelain *et al.*, Safe harbours for the integration of new DNA in the human genome. Nat Rev Cancer 12, 51-58 (2012); the entire contents of each of which are incorporated herein by reference). Five reporters and corresponding guide RNA vector pairs containing sequences identical to those in the genome were constructed. To evaluate the orthogonality of recCas9 when programmed with different guide RNAs, all combinations of five guide RNA pairs with five reporters were tested.

[00183] Cotransfection of reporter, guide RNA plasmids, and recCas9 expression vectors revealed that three of the five reporters tested resulted in substantial levels of EGFP-positive cells consistent with recCas9-mediated recombination. This *EGFP* expression was strictly dependent upon cotransfection with a recCas9 expression vector and guide RNA plasmids matching the target site sequences on the reporter construct (Figure 4A). The same

guide RNA pairs that caused recombination when cotransfected with cognate reporter plasmids and a recCas9 vector were unable to mediate recombination when cotransfected with non-cognate reporter plasmids (Figure 4A). These results demonstrate that recCas9 activity is orthogonal and will only catalyze recombination at a gix related core sites when programmed with a pair of guide RNAs matching the flanking sequences. No recombinase activity above the background level of the assay was observed when reporter plasmids were transfected without vectors expressing recCas9 and guide RNAs.

Characterization of recCas9 products

[00184] The products of recCas9-mediated recombination of the reporter plasmids were characterized to confirm that *EGFP* expression was a result of recCas9-mediated removal of the poly-A terminator sequence. Reporter plasmids were sequenced for chromosome 5-site 1, chromosome 12, and chromosome 13 (*FGF14* locus) after cotransfection with recCas9 expression vectors and with plasmids producing cognate or non-cognate guide RNA pairs. After incubation for 72 hours, episomal DNA was extracted (as described above) and transformed into *E. coli* to isolate reporter plasmids. Single colonies containing reporter plasmids were sequenced (Figure 4B).

[00185] Individual colonies were expected to contain either an unmodified or a recombined reporter plasmid (Figure 4C). For each biological replicate, an average of 97 colonies transformed with reporter plasmid isolated from each transfection condition were sequenced. Recombined plasmids were only observed if reporter plasmids were previously cotransfected with cognate guide RNA plasmids and recCas9 expression vectors (Figure 4D). In two separate experiments, the percent of recombined plasmid ranged from 12% for site 1 in chromosome 5 to an average of 32% for the *FGF14* locus in chromosome 13. The sequencing data therefore were consistent with the earlier flow cytometry analysis in Figure 4A. The absolute levels of recombined plasmid were somewhat higher than the percent of EGFP-positive cells (Figure 4). This difference likely arises because the flow cytometry assay does not report on multiple recombination events that can occur when multiple copies of the reporter plasmid are present in a single cell; even a single recombination event may result in EGFP fluorescence. As a result, the percentage of EGFP-positive cells may correspond to a lower limit on the actual percentage of recombined reporter plasmids. Alternatively, the difference may reflect the negative correlation between plasmid size and transformation efficiency (see, *e.g.*, Hanahan, Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166, 557-580 (1983), the entire contents of which is hereby

incorporated by reference); the recombined plasmid is approximately 5,700 base pairs and may transform slightly better than the intact plasmid, which is approximately 6,900 base pairs.

[00186] Since zinc finger-recombinases have been reported to cause mutations at recombinase core-site junctions (see, *e.g.*, Gaj *et al.*, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. Nucleic acids research 41, 3937-3946 (2013), the entire contents of which is hereby incorporated by reference), whether such mutagenesis occurs from recCas9 treatment was tested. In the reporter construct, recCas9 should delete *kanR* and the poly-A terminator by first cleaving the central dinucleotide of both gix core sites and then religating the two cores to each other (Figure 4C). Thus, the recombination product should be a single recombination site consisting of the first half of the 'left' target site and the second half of the 'right' target site. Erroneous or incomplete reactions could result in other products. Strikingly, all of the 134 recombined sequences examined contained the expected recombination products. Further, a total of 2,317 sequencing reads from two separate sets of transfection experiments revealed only three sequencing reads containing potential deletion products at otherwise non-recombined plasmids.

[00187] One of these deletion-containing reads was observed in a chromosome 12 reporter plasmid that was transfected with the pUC control and lacked both recCas9 target sites as well as the polyA terminator. This product was attributed to DNA damage that occurred during the transfection, isolation, or subsequent manipulation. Because recCas9 may only localize to sequences when cotransfected with reporter and cognate guide RNA expression vectors, a more relevant metric may be to measure the total number of deletion products observed when reporter plasmids are cotransfected with cognate guide RNA vectors and recCas9 expression vectors. A single indel was observed out of a total of 185 plasmids sequenced from cotransfections with the chromosome 5-site 1 reporter and cognate guide RNA. Similarly, one indel was observed out of 204 plasmids from the chromosome 12 reporter following transfection with cognate guide RNA and recCas9 expression vectors. Notably, out of 202 sequencing reads, no indels were observed from the chromosome 13 reporter following cognate guide RNA and recCas9 cotransfection, despite resulting in the highest observed levels of recombination. These observations collectively suggest that recCas9 mediates predominantly error-free recombination.

[00188] Taken together, these results establish that recCas9 can target multiple sites found within the human genome with minimal cross-reactivity or byproduct formation.

Substrates undergo efficient recombination only in the presence of cognate guide RNA sequences and recCas9, give clean recombination products in human cells, and generally do not result in mutations at the core-site junctions or products such as indels that arise from cellular DNA repair.

RecCas9-mediated genomic deletion

[00189] Finally, whether recCas9 is capable of operating directly on the genomic DNA of cultured human cells was investigated. Using the list of potential recCas9 recognition sites in the human genome (Table 9), pairs of sites that, if targeted by recCas9, would yield chromosomal deletion events detectable by PCR, were sought. Guide RNA expression vectors were designed to direct recCas9 to those recCas9 sites closest to the chromosome 5-site 1 or chromosome 13 (*FGF14 locus*), sites which were both shown to be recombined in transient transfection assays (Figure 4). The new target sites ranged from approximately 3 to 23 Mbp upstream and 7 to 10 Mbp downstream of chromosome 5-site 1, and 12 to 44 Mbp upstream of the chromosome 13-FGF14 site. The recCas9 expression vector was cotransfected with each of these new guide RNA pairs and the validated guide RNA pairs used for chromosome 5-site 1 or chromosome 13-FGF14, but evidence of chromosomal deletions by genomic PCR was not observed.

[00190] It was thought that genomic deletion might be more efficient if the recCas9 target sites were closer to each other on the genome. Two recCas9 sites separated by 14.2 kb within an intronic region of *FAM19A2* were identified; these sites also contained identical dinucleotide cores which should facilitate deletion. *FAM19A2* is one of five closely related TAFA-family genes encoding small, secreted proteins that are thought to have a regulatory role in immune and nerve cells (see, *e.g.*, Parker *et al.*, Admixture mapping identifies a quantitative trait locus associated with FEV1/FVC in the COPDGene Study. Genet Epidemiol 38, 652-659 (2014), the entire contents of which is hereby incorporated by reference). Small nucleotide polymorphisms located in intronic sequences of *FAM19A2* have been associated with elevated risk for systemic lupus erythematosus (SLE) and chronic obstructive pulmonary disease (COPD) in genome-wide association studies (see, *e.g.*, Parker *et al.*, Admixture mapping identifies a quantitative trait locus associated with FEV1/FVC in the COPDGene Study. Genet Epidemiol 38, 652-659 (2014), the entire contents of which is hereby incorporated by reference); deletion of the intronic regions of this gene might therefore provide insights into the causes of these diseases. Four guide RNA sequences were cloned in expression vectors designed to mediate recCas9 deletion between these two

FAM19A2 sites. Vectors expressing these guide RNAs were cotransfected with the recCas9 expression vector (Figure 5A). RecCas9-mediated recombination between the two sites should result in deletion of the 14.2 kb intervening region. Indeed, this deletion event was detected by nested PCR using gene-specific primers that flank the two *FAM19A2* recCas9 targets. The expected PCR product that is consistent with recCas9-mediated deletion was observed only in genomic DNA isolated from cells cotransfected with the recCas9 and all four guide RNA expression vectors (Figure 5B). The deletion PCR product was not detected in the genomic DNA of cells transfected without either the upstream or downstream pair of guide RNA expression vectors alone, without the recCas9 expression plasmid, or for the genomic DNA of untransfected control cells (Figure 5B). The estimated limit of detection for these nested PCR products was approximately 1 deletion event per 5,500 chromosomal copies. The 415-bp PCR product corresponding to the predicted genomic deletion was isolated and sequenced. Sequencing confirmed that the PCR product matched the predicted junction expected from the recombinase-mediated genomic deletion and did not contain any insertions or deletions suggestive of NHEJ (Figure 5C).

[00191] A lower limit on the minimum genomic deletion efficiency was estimated using nested PCR on the serial dilutions of genomic template (see above or, *e.g.*, Sykes *et al.*, Quantitation of targets for PCR by use of limiting dilution. Biotechniques 13, 444-449 (1992), the entire contents of which is hereby incorporated by reference, for greater detail). A given amount of genomic DNA that yields the recCas9-specific nested PCR product must contain at least one edited chromosome. To establish a lower limit on this recCas9-mediated genomic deletion event, nested PCR was performed on serial dilutions of genomic DNA (isolated from cells transfected with recCas9 and the four FAM19A2 guide RNA expression vectors) to determine the lowest concentration of genomic template DNA that results in a detectable deletion product. These experiments revealed a lower limit of deletion efficiency of 0.023±0.017% (average of three biological replicates) (Figure 5D), suggesting that recCas9-mediated genomic deletion proceeds with at least this efficiency. Nested PCR of the genomic DNA of untransfected cells resulted in no product, with an estimated limit of detection of < 0.0072% recombination.

Use of other alternative recombinases

[00192] A Cre recombinase evolved to target a site in the Rosa locus of the human genome called "36C6" was fused to to dCas9. This fusion was then used to recombine a

plasmid-based reporter containing the Rosa target site in a guide-RNA dependent fashion. Figure 7A demonstrates the results of linker optimization using wild-type Cre and 36C6. The 1x 2x, 5x, and 8x linkers shown are the number of GGS repeats in the linker. Reversion analysis demonstrated that making mutations to 36C6 fused to dCas9 could impact the relative guide dependence of the chimeric fusion (Figure 7B). Reversions are labeled with their non-mutated amino acids. For example, position 306, which had been mutated to an M, was reverted to an I before the assay was performed. A GinB construct, targeting its cognate reporter, was used as a control for the experimental data shown in Figures 7A and 7B. The on-target guides were the chr13-102010574 guides (plasmids BC165 and 166). Abbreviations shown are GGS-36C6: dCas9-GGS-36C6; 2GGS-36C6 (using linker SEQ ID NO: 182): sdCas9-GGSGGS-36C6 (using linker SEQ ID NO: 182).

[00193] The target sequence used for 36C6 and all variant transfections is shown below: (guides – italics; Rosa site – bold):
CCT*AGGGAAGTGATCATAGCTGA*GTTTCT**ATCTCATGGTTTATGCTAAACTATAT GTTGACAT**GTTGA*GGAGACTTAAGTCCAAAACC*TGG (SEQ ID NO: 760)

[00194] In **Figures 7A, 7B, 8, 9A, and 9B**, the on-target guides for GinB were the chr13-102010574 guides (plasmids BC165 and 166). All off-target guides in **Figures 7A, 7B, 8, 9A, and 9B** were composed of the chr12-62418577 guides (BC163 and BC164).

[00195] PAMs were identified flanking the Rosa26 site in the human genome that could support dCas9 binding (Figure 8, top). Guide RNAs and a plasmid reporter were then designed to test whether the endogenous protospacers could support dCas9-36C6 activity. A GinB construct, targeting its cognate reporter, was used as a control. See Figure 8. Mix: equal parts mixture of all 5 linker variants between Cas9 and 36C6. For hRosa, the target sequence, including guide RNA tagets, are below: (guides – italics; Rosa site – bold)
CCT*GAAATAATGCAAGTGTAGAA*TAACTTTTTAAA**ATCTCATGGTTTATGCTAAAC TATATGTTGACAT**AAG*AGTGGTGATAAGGCAACAGT*AGG (SEQ ID NO: 767)

[00196] The on target guide plasmids for hRosa are identical to the other gRNA expression plasmids, except the protospacers are replaced with those shown above (Figure 8).

[00197] Several tested Cre truncations of dCas9-Cre recombinase fusions are shown in **Figure 9A**. Truncated variants of Cre recombinase fused to dCas9 showed both appreciable recombinase activity as well as a strict reliance on the presence of guide RNA in a Lox

plasmid reporter system (Figure 9B). Truncated variants are labeled with the residue at which the truncated Cre begins. The linker for all fusion proteins shown in Figures 9A and 9B is 8xGGS. Wild type Cre fused to dCas9 was used as a positive control. The target sequence used for 36C6 and all variant transfections is shown below: (guides – italics; Rosa site – bold):

CCT*AGGGAAGTGATCATAGCTGA*GTTTCT**ATCTCATGGTTTATGCTAAACTATAT GTTGACAT**GTTGA*GGAGACTTAAGTCCAAAACC*TGG (SEQ ID NO: 768)

[00198] The on-target guides used were the chr13-102010574 guides (plasmids BC165 and 166) and the off-target guides were the chr12-62418577 guide (BC163 and BC164).

References

1. J. A. Doudna, E. Charpentier, Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
2. M. R. Capecchi, Altering the genome by homologous recombination. *Science* **244**, 1288-1292 (1989).
3. K. R. Thomas, K. R. Folger, M. R. Capecchi, High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419-428 (1986).
4. A. Choulika, A. Perrin, B. Dujon, J. F. Nicolas, Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol Cell Biol* **15**, 1968-1973 (1995).
5. D. Carroll, Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther* **15**, 1463-1468 (2008).
6. J. C. Miller *et al.*, A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* **29**, 143-U149 (2011).
7. J. K. Joung, J. D. Sander, TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**, 49-55 (2013).
8. P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).
9. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
10. J. P. Guilinger, D. B. Thompson, D. R. Liu, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology*, (2014).

11. S. Q. Tsai *et al.*, Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology*, (2014).

12. H. Fung, D. M. Weinstock, Repair at single targeted DNA double-strand breaks in pluripotent and differentiated human cells. *PloS one* **6**, e20514 (2011).

13. W. D. Heyer, K. T. Ehmsen, J. Liu, Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**, 113-139 (2010).

14. D. Branzei, M. Foiani, Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Bio* **9**, 297-308 (2008).

15. V. T. Chu *et al.*, Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology*, (2015).

16. T. Maruyama *et al.*, Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology*, (2015).

17. S. Lin, B. T. Staahl, R. K. Alla, J. A. Doudna, Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).

18. S. Turan, C. Zehe, J. Kuehle, J. H. Qiao, J. Bode, Recombinase-mediated cassette exchange (RMCE) - A rapidly-expanding toolbox for targeted genomic modifications. *Gene* **515**, 1-27 (2013).

19. T. Gaj, S. J. Sirk, C. F. Barbas, Expanding the Scope of Site-Specific Recombinases for Genetic and Metabolic Engineering. *Biotechnology and bioengineering* **111**, 1-15 (2014).

20. N. D. F. Grindley, K. L. Whiteson, P. A. Rice, Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**, 567-605 (2006).

21. C. R. Sclimenti, B. Thyagarajan, M. P. Calos, Directed evolution of a recombinase for improved genomic integration at a native human sequence. *Nucleic acids research* **29**, 5044-5051 (2001).

22. R. Shah, F. Li, E. Voziyanova, Y. Voziyanov, Target-specific variants of Flp recombinase mediate genome engineering reactions in mammalian cells. *The FEBS journal* **282**, 3323-3333 (2015).

23. J. Karpinski *et al.*, Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nature biotechnology*, (2016).

24. F. Buchholz, A. F. Stewart, Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nature biotechnology* **19**, 1047-1052 (2001).

25. B. Thyagarajan, E. C. Olivares, R. P. Hollis, D. S. Ginsburg, M. P. Calos, Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol* **21**, 3926-3934 (2001).

26. B. Thyagarajan, M. J. Guimaraes, A. C. Groth, M. P. Calos, Mammalian genomes contain active recombinase recognition sites. *Gene* **244**, 47-54 (2000).

27. A. Akopian, J. He, M. R. Boocock, W. M. Stark, Chimeric recombinases with designed DNA sequence recognition. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8688-8691 (2003).

28. R. M. Gordley, C. A. Gersbach, C. F. Barbas, 3rd, Synthesis of programmable integrases. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5053-5058 (2009).

29. M. M. Prorocic *et al.*, Zinc-finger recombinase activities in vitro. *Nucleic acids research* **39**, 9316-9328 (2011).

30. C. A. Gersbach, T. Gaj, R. M. Gordley, A. C. Mercer, C. F. Barbas, Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. *Nucleic acids research* **39**, 7868-7878 (2011).

31. A. C. Mercer, T. Gaj, R. P. Fuller, C. F. Barbas, Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic acids research* **40**, 11163-11172 (2012).

32. T. Matsuda, C. L. Cepko, Controlled expression of transgenes introduced by in vivo electroporation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1027-1032 (2007).

33. N. E. Sanjana *et al.*, A transcription activator-like effector toolbox for genome engineering. *Nature protocols* 7, 171-192 (2012).

34. T. Gaj, A. C. Mercer, S. J. Sirk, H. L. Smith, C. F. Barbas, A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic acids research* **41**, 3937-3946 (2013).

35. Y. Fu *et al.*, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826 (2013).

36. J. Quan, J. Tian, Circular polymerase extension cloning of complex gene libraries and pathways. *PloS one* **4**, e6441 (2009).

37. N. Hillson. (2010), vol. 2015, pp. CPEC protocol.

38. R. C. Gentleman *et al.*, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).

39. K. Motmans, S. Thirion, J. Raus, C. Vandevyver, Isolation and quantification of episomal expression vectors in human T cells. *Biotechniques* **23**, 1044-1046 (1997).

40. B. Hirt, Selective extraction of polyoma DNA from infected mouse cell cultures. *J Mol Biol* **26**, 365-369 (1967).

41. A. Klippel, G. Mertens, T. Patschinsky, R. Kahmann, The DNA Invertase Gin of Phage Mu - Formation of a Covalent Complex with DNA Via a Phosphoserine at Amino-Acid Position-9. *Embo Journal* 7, 1229-1237 (1988).

42. G. Mertens *et al.*, Site-specific recombination in bacteriophage Mu: characterization of binding sites for the DNA invertase Gin. *The EMBO journal* 7, 1219-1227 (1988).

43. R. H. Plasterk, A. Brinkman, P. van de Putte, DNA inversions in the chromosome of Escherichia coli and in bacteriophage Mu: relationship to other site-specific recombination systems. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 5355-5358 (1983).

44. H. Xu *et al.*, Sequence determinants of improved CRISPR sgRNA design. *Genome research* **25**, 1147-1157 (2015).

45. R. M. Gordley, J. D. Smith, T. Graslund, C. F. Barbas, 3rd, Evolution of programmable zinc finger-recombinases with activity in human cells. *J Mol Biol* **367**, 802-813 (2007).

46. C. A. Gersbach, T. Gaj, R. M. Gordley, C. F. Barbas, 3rd, Directed evolution of recombinase specificity by split gene reassembly. *Nucleic acids research* **38**, 4198-4206 (2010).

47. T. Gaj, A. C. Mercer, C. A. Gersbach, R. M. Gordley, C. F. Barbas, Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 498-503 (2011).

48. T. Gaj *et al.*, Enhancing the Specificity of Recombinase-Mediated Genome Engineering through Dimer Interface Redesign. *J Am Chem Soc* **136**, 5047-5056 (2014).

49. F. Cunningham *et al.*, Ensembl 2015. *Nucleic acids research* **43**, D662-669 (2015).

50. J. C. van Swieten *et al.*, A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected]. *Am J Hum Genet* **72**, 191-199 (2003).

51. E. Brusse *et al.*, Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14 gene (SCA27): A new phenotype. *Mov Disord* **21**, 396-401 (2006).

52. K. Choquet, R. La Piana, B. Brais, A novel frameshift mutation in FGF14 causes an autosomal dominant episodic ataxia. *Neurogenetics* **16**, 233-236 (2015).

53. J. A. Coebergh *et al.*, A new variable phenotype in spinocerebellar ataxia 27 (SCA 27) caused by a deletion in the FGF14 gene. *Eur J Paediatr Neurol* **18**, 413-415 (2014).

54. K. Shimojima *et al.*, Spinocerebellar ataxias type 27 derived from a disruption of the fibroblast growth factor 14 gene with mimicking phenotype of paroxysmal non-kinesigenic dyskinesia. *Brain Dev* **34**, 230-233 (2012).

55. M. Sadelain, E. P. Papapetrou, F. D. Bushman, Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer* **12**, 51-58 (2012).

56. D. Hanahan, Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166**, 557-580 (1983).

57. M. M. Parker *et al.*, Admixture mapping identifies a quantitative trait locus associated with FEV1/FVC in the COPDGene Study. *Genet Epidemiol* **38**, 652-659 (2014).

58. P. J. Sykes *et al.*, Quantitation of targets for PCR by use of limiting dilution. *Biotechniques* **13**, 444-449 (1992).

59. A. Rath, R. Hromas, A. De Benedetti, Fidelity of end joining in mammalian episomes and the impact of Metnase on joint processing. *BMC Mol Biol* **15**, 6 (2014).

60. P. Rebuzzini *et al.*, New mammalian cellular systems to study mutations introduced at the break site by non-homologous end-joining. *DNA Repair (Amst)* **4**, 546-555 (2005).

61. J. Smith, C. Baldeyron, I. De Oliveira, M. Sala-Trepat, D. Papadopoulo, The influence of DNA double-strand break structure on end-joining in human cells. *Nucleic acids research* **29**, 4783-4792 (2001).

62. S. Turan *et al.*, Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J Mol Biol* **407**, 193-221 (2011).

63. S. J. Sirk, T. Gaj, A. Jonsson, A. C. Mercer, C. F. Barbas, Expanding the zinc-finger recombinase repertoire: directed evolution and mutational analysis of serine recombinase specificity determinants. *Nucleic acids research* **42**, 4755-4766 (2014).

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

65. B. P. Kleinstiver *et al.*, Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-U249 (2015).

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

67. B. Zetsche *et al.*, Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759-771 (2015).

68. K. Dormiani *et al.*, Long-term and efficient expression of human beta-globin gene in a hematopoietic cell line using a new site-specific integrating non-viral system. *Gene Ther* **22**, 663-674 (2015).

69. E. Wijnker, H. de Jong, Managing meiotic recombination in plant breeding. *Trends in plant science* **13**, 640-646 (2008).

70. J. F. Petolino, V. Srivastava, H. Daniell, Editing Plant Genomes: a new era of crop improvement. *Plant Biotechnol J* **14**, 435-436 (2016).

EQUIVALENTS AND SCOPE

[00199] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[00200] In the claims articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention also includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

[00201] Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, *etc.*, from one or more of the claims or from relevant portions of

the description is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for any of the purposes disclosed herein are included, and methods of making the composition according to any of the methods of making disclosed herein or other methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[00202] Where elements are presented as lists, *e.g.,* in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It is also noted that the term "comprising" is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, steps, *etc.*, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, steps, *etc.* For purposes of simplicity those embodiments have not been specifically set forth *in haec verba* herein. Thus for each embodiment of the invention that comprises one or more elements, features, steps, *etc.*, the invention also provides embodiments that consist or consist essentially of those elements, features, steps, *etc.*

[00203] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[00204] In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims.

For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

[00205] All publications, patents and sequence database entries mentioned herein, including those items listed above, are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

CLAIMS

What is claimed is:

1. A fusion protein comprising:
(i) a guide nucleotide sequence-programmable DNA binding protein domain;
(ii) a linker; and
(iii) a recombinase catalytic domain.

2. The fusion protein of claim 1, wherein the guide nucleotide sequence-programmable DNA binding protein domain is selected from the group consisting of nuclease inactive Cas9 (dCas9) domains, nuclease inactive Cpf1 domains, nuclease inactive Argonaute domains, and variants thereof.

3. The fusion protein of claim 1, wherein the guide nucleotide sequence-programmable DNA-binding protein domain is a nuclease inactive Cas9 (dCas9) domain.

4. The fusion protein of claim 2 or 3, wherein the amino acid sequence of the dCas9 domain comprises mutations corresponding to D10A or H840A mutation in SEQ ID NO: 1.

5. The fusion protein of any one of claims 2-4, wherein the amino acid sequence of the dCas9 domain comprises a mutation corresponding to a D10A mutation in SEQ ID NO: 1 and a mutation corresponding to an H840A mutation in SEQ ID NO: 1.

6. The fusion protein of any one of claims 2-5, wherein the amino acid sequence of the dCas9 domain further comprises a mutation corresponding to a missing N-terminal methionine in SEQ ID NO: 1.

7. The fusion protein of claim 2 or 3, wherein the amino acid sequence of the dCas9 domain comprises SEQ ID NO: 712.

8. The fusion protein of claim 2 or 3, wherein the amino acid sequence of the dCas9 domain has 95% or greater sequence indentity with SEQ ID NO: 712.

9. The fusion protein of claim 2 or 3, wherein the amino acid sequence of the dCas9 domain has 96%, 97%, 98%, 99%, or greater sequence indentity with SEQ ID NO: 712.

10. The fusion protein of any preceding claim, wherein the recombinase catalytic domain is a serine recombinase catalytic domain or a tyrosine recombinase catalytic domain.

11. The fusion protein of claim 10, wherein the serine recombinase catalytic domain or tyrosine recombinase catalytic domain is selected from a Gin, Sin, Tn3, Hin, β, γδ, PhiC31, Cre, or FLP recombinase catalytic domain.

12. The fusion protein of claim 10 or claim 11, wherein the amino acid sequence of the Gin recombinase catalytic domain comprises one or more mutations selected from the group consisting of: an H106Y, I127L, I136R, or G137F mutation in SEQ ID NO: 713.

13. The fusion protein of claim 11 or claim 12, wherein the amino acid sequence of the Gin recombinase catalytic domain comprises SEQ ID NO: 713.

14. The fusion protein of claim 10 or claim 11, wherein the amino acid sequence of the Cre recombinase is truncated.

15. The fusion protein of claim 11 or claim 14, wherein the tyrosine recombinase catalytic domain is the 25 kDa carboxy-terminal domain of the Cre recombinase.

16. The fusion protein of claim 14, wherein the Cre recombinase begins with amino acid R118, A127, E138, or R154.

17. The fusion protein of any one of claims 1-16, wherein the amino acid sequence of the recombinase has been further mutated.

18. The fusion protein of any preceding claim, wherein the recombinase catalytic domain is an evolved recombinase catalytic domain.

19. The fusion protein of any one of claims 1-18, wherein the linker has a length of about 33 angstroms to about 81 angstroms.

20. The fusion protein of any one of claims 1-19, wherein the linker is a peptide linker.

21. The fusion protein of claim 20, wherein the peptide linker comprises an XTEN linker (SGSETPGTSESATPES (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9)), an amino acid sequence comprising one or more repeats of the tri-peptide GGS, or any of the following amino acid sequences: VPFLLEPDNINGKTC (SEQ ID NO: 10), GSAGSAAGSGEF (SEQ ID NO: 11), SIVAQLSRPDPA (SEQ ID NO: 12), MKIIEQLPSA (SEQ ID NO: 13), VRHKLKRVGS (SEQ ID NO: 14), GHGTGSTGSGSS (SEQ ID NO: 15), MSRPDPA (SEQ ID NO: 16), or GGSM (SEQ ID NO: 17).

22. The fusion protein of claim 20 or claim 21, wherein the peptide linker comprises one or more repeats of the tri-peptide GGS.

23. The fusion protein of any one of claims 20-22, wherein the peptide linker comprises from six to ten repeats of the tri-peptide GGS.

24. The fusion protein of any one of claims 20-23, wherein the peptide linker comprises eight repeats of the tri-peptide GGS.

25. The fusion protein of any one of claims 20-24, wherein the peptide linker is from about 18 to about 27 amino acids long.

26. The fusion protein of any one of claims 20-25, wherein the peptide linker is 24 amino acids long.

27. The fusion protein of any one of claims 20-26, wherein the peptide linker has the amino acid sequence GGSGGSGGSGGSGGSGGSGGSGGS (SEQ ID NO: 183).

28. The fusion protein of any one of claims 1-19, wherein the linker is a non-peptide linker.

29. The fusion protein of claim 28, wherein the non-peptide linker comprises polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacryl amide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker.

30. The fusion protein of claim 29, wherein the alkyl linker has the formula —NH—(CH2)s—C(O)—, wherein s may be any integer from 1-100, inclusive.

31. The fusion protein of claim 30, wherein s is an integer from 1-20, inclusive.

32. The fusion protein of any one of claims 1-31, further comprising a nuclear localization signal (NLS) domain.

33. The fusion protein of claim 32, wherein the NLS domain is bound to the guide nucleotide sequence-programmable DNA binding protein domain or the Gin recombinase catalytic domain via one or more second linkers.

34. The fusion protein of claim 33, wherein the second linker is a peptide linker.

35. The fusion protein of claim 33 or claim 34, wherein the second linker comprises an XTEN linker (SGSETPGTSESATPES (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9)), an amino acid sequence comprising one or more repeats of the tri-peptide GGS, or any of the following amino acid sequences: VPFLLEPDNINGKTC (SEQ ID NO: 10), GSAGSAAGSGEF (SEQ ID NO: 11), SIVAQLSRPDPA (SEQ ID NO: 12), MKIIEQLPSA (SEQ ID NO: 13), VRHKLKRVGS (SEQ ID NO: 14), GHGTGSTGSGSS (SEQ ID NO: 15), MSRPDPA (SEQ ID NO: 16), or GGSM (SEQ ID NO: 17).

36. The fusion protein of any one of claims 33-35, wherein the second linker comprises one or more repeats of the tri-peptide GGS.

37. The fusion protein of any one of claims 33-36, wherein the second linker comprises from one to five repeats of the tri-peptide GGS.

38. The fusion protein of any one of claims 33-37, wherein the second linker comprises one repeat of the tri-peptide GGS.

39. The fusion protein of any one of claims 33-38, wherein the second linker has the sequence GGS.

40. The fusion protein of claim 33, wherein the second linker is a non-peptide linker.

41. The fusion protein of claim 40, wherein the non-peptide linker comprises polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacryl amide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker.

42. The fusion protein of claim 41, wherein the alkyl linker has the formula —NH—(CH2)s—C(O)—, wherein s may be any integer from 1-100, inclusive.

43. The fusion protein of claim 42, wherein s is any integer from 1-20, inclusive.

44. The fusion protein of any one of claims 32-43, wherein the fusion protein comprises the structure NH_2-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH.

45. The fusion protein of any one of claims 32-44, wherein the fusion protein comprises the structure NH_2-[Gin recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH.

46. The fusion protein of any one of claims 32-39, 44, or 45, wherein the fusion protein comprises the amino acid sequence shown in SEQ ID NO: 719.

47. The fusion protein of any preceding claim, further comprising one or more affinity tags.

48. The fusion protein of claim 47, wherein the affinity tag is selected from the group consisting of a FLAG tag, a polyhistidine (poly-His) tag, a polyarginine (poly-Arg) tag, a Myc tag, and an HA tag.

49. The fusion protein of claim 47 or claim 48, wherein the affinity tag is a FLAG tag.

50. The fusion protein of any one of claims 47-49, wherein the FLAG tag has the sequence PKKKRKV (SEQ ID NO: 702).

51. The fusion protein of any one of claims 47-50, wherein the one or more affinity tags are bound to the guide nucleotide sequence-programmable DNA binding protein domain, the recombinase catalytic domain, or the NLS domain via one or more third linkers.

52. The fusion protein of claim 51, wherein the third linker is a peptide linker.

53. The fusion protein of claim 51 or claim 52, wherein the third linker comprises an XTEN linker, (SGSETPGTSESATPES (SEQ ID NO: 7), SGSETPGTSESA (SEQ ID NO: 8), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 9)), an amino acid sequence comprising one or more repeats of the tri-peptide GGS, or any of the following amino acid sequences: VPFLLEPDNINGKTC (SEQ ID NO: 10), GSAGSAAGSGEF (SEQ ID NO: 11), SIVAQLSRPDPA (SEQ ID NO: 12), MKIIEQLPSA (SEQ ID NO: 13), VRHKLKRVGS (SEQ ID NO: 14), GHGTGSTGSGSS (SEQ ID NO: 15), MSRPDPA (SEQ ID NO: 16), or GGSM (SEQ ID NO: 17).

54. The fusion protein of any one of claims 51-53, wherein the third linker comprises one or more repeats of the tri-peptide GGS.

55. The fusion protein of any one of claims 51-54, wherein the third linker comprises from one to five repeats of the tri-peptide GGS.

56. The fusion protein of any one of claims 51-55, wherein the third linker comprises one repeat of the tri-peptide GGS.

57. The fusion protein of any one of claims 51-56, wherein the third linker has the sequence GGS.

58. The fusion protein of claim 51, wherein the third linker is a non-peptide linker.

59. The fusion protein of claim 58, wherein the non-peptide linker comprises polyethylene glycol (PEG), polypropylene glycol (PPG), co-poly(ethylene/propylene) glycol, polyoxyethylene (POE), polyurethane, polyphosphazene, polysaccharides, dextran, polyvinyl alcohol, polyvinylpyrrolidones, polyvinyl ethyl ether, polyacryl amide, polyacrylate, polycyanoacrylates, lipid polymers, chitins, hyaluronic acid, heparin, or an alkyl linker.

60. The fusion protein of claim 59, wherein the alkyl linker has the formula —NH—(CH2)s—C(O)—, wherein s may be any integer from 1-100, inclusive.

61. The fusion protein of claim 60, wherein s may be any integer from 1-20.

62. The fusion protein of any one of claims 47-61, wherein the fusion protein comprises the structure NH_2-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[affinity tag]-COOH, NH_2-[guide nucleotide sequence-programmable DNA binding protein domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH, NH_2-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-[optional linker sequence]-[optional affinity tag]-COOH, NH_2-[affinity tag]-[optional linker sequence]-[recombinase catalytic domain]-[linker sequence]-[guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH, NH_2-[affinity tag]-[optional linker sequence]-[guide nucleotide sequence-programmable DNA binding protein

domain]-[linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[NLS domain]-COOH, or NH_2-[affinity tag]-[optional linker sequence]-[N-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[recombinase catalytic domain]-[optional linker sequence]-[C-terminal portion of a bifurcated or circularly permuted guide nucleotide sequence-programmable DNA binding protein domain]-[optional linker sequence]-[NLS domain]-COOH.

63. The fusion protein of any one of claims 47-57, wherein the fusion protein has greater than 99% sequence identity with the amino acid sequence shown in SEQ ID NO: 185.

64. The fusion protein of any one of claims 47-57, wherein the fusion protein has greater than 90% or 95% sequence identity with the amino acid sequence shown in SEQ ID NO: 185.

65. The fusion protein of any one of claims 47-57, 63, or 64, wherein the fusion protein has the amino acid sequence shown in SEQ ID NO: 185.

66. The fusion protein of any preceding claim, wherein the guide nucleotide sequence-programmable DNA binding protein domain is bound to a guide RNA (gRNA).

67. A dimer of the fusion protein of claim 66.

68. The dimer of the fusion protein of claim 67, wherein the dimer is bound to a DNA molecule.

69. The dimer of the fusion protein of claim 68, wherein each fusion protein of the dimer is bound to the same strand of the DNA molecule.

70. The dimer of the fusion protein of claim 68, wherein each fusion protein of the dimer is bound to opposite strands of the DNA molecule.

71. The dimer of the fusion protein of any one of claims 67-70, wherein the gRNAs of the dimer hybridize to gRNA binding sites flanking a recombinase site.

72. The method of claim 71, wherein the recombinase site comprises a res, gix, hix, six, resH, LoxP, FTR, or att core or related core sequence.

73. The dimer of the fusion protein of claim 71 or claim 72, wherein the recombinase site comprises a gix core or gix-related core sequence.

74. The dimer of claim 72 or claim 73, wherein the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 3 to 7 base pairs.

75. The dimer of any one of claims 72-74, wherein the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 5 to 6 base pairs.

76. The dimer of the fusion protein of any one of claims 67-75, wherein a first dimer binds to a second dimer thereby forming a tetramer of the fusion protein.

77. A tetramer of the fusion protein of claim 66.

78. The tetramer of the fusion protein of claim 77, wherein the tetramer is bound to a DNA molecule.

79. A tetramer of the fusion of claim 68, wherein each dimer is bound to an opposite strand of DNA.

80. A tetramer of the fusion protein of claim 68, wherein each dimer is bound to the same strand of DNA.

81. A method for site-specific recombination between two DNA molecules, comprising:

(a) contacting a first DNA with a first fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain binds a first gRNA that hybridizes to a first region of the first DNA;

(b) contacting the first DNA with a second fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the first DNA;

(c) contacting a second DNA with a third fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the third fusion protein binds a third gRNA that hybridizes to a first region of the second DNA; and

(d) contacting the second DNA with a fourth fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a second region of the second DNA;

wherein the binding of the fusion proteins in steps (a)-(d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNAs are recombined, and wherein the first, second, third, and/or fourth fusion protein is a fusion protein of any one of claims 1-63.

82. The method of claim 81, wherein the first and second DNA molecules have different sequences.

83. The method of claim 81, wherein the gRNAs of steps (a) and (b) hybridize to opposing strands of the first DNA, and the gRNAs of steps (c) and (d) hybridize to opposing strands of the second DNA.

84. The method of any one of claims 81-83, wherein the gRNAs of steps (a) and (b); and/or the gRNAs of steps (c) and (d) hybridize to regions of their respective DNAs that are no more than 100 base pairs apart.

85. The method of claim 84, wherein the gRNAs of steps (a) and (b); and/or the gRNAs of steps (c) and (d) hybridize to regions of their respective DNAs that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, or no more than 90 base pairs apart.

86. The method of any one of claims 81-85, wherein the gRNAs of steps (a) and (b); and/or the gRNAs of steps (c) and (d) hybridize to regions of their respective DNAs at gRNA binding sites that flank a recombinase site.

87. The method of claim 86, wherein the recombinase site comprises a res, gix, hix, six, resH, LoxP, FTR, or att core or related core sequence.

88. The method of claim 86 or claim 87, wherein the recombinase site comprises a gix core or gix-related core sequence.

89. The method of claim 87 or claim 88, wherein the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 3 to 7 base pairs.

90. The method of any one of claims 87-89, wherein the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 5 to 6 base pairs.

91. A method for site-specific recombination between two regions of a single DNA molecule, comprising:

(a) contacting the DNA with a first fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain binds a first gRNA that hybridizes to a first region of the DNA;

(b) contacting the DNA with a second fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the second fusion protein binds a second gRNA that hybridizes to a second region of the DNA;

(c) contacting the DNA with a third fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the third fusion protein binds a third gRNA that hybridizes to a third region of the DNA; and

(d) contacting the DNA with a fourth fusion protein, wherein the guide nucleotide sequence-programmable DNA binding protein domain of the fourth fusion protein binds a fourth gRNA that hybridizes to a fourth region of the DNA;

wherein the binding of the fusion proteins in steps (a)-(d) results in the tetramerization of the recombinase catalytic domains of the fusion proteins, under conditions such that the DNA is recombined, and wherein the first, second, third, and/or fourth fusion protein is the fusion protein of any one of claims 1-63.

92. The method of claim 91, wherein the two regions of the single DNA molecule that are recombined have different sequences.

93. The method of claim 91 or claim 92, wherein the recombination results in the deletion of a region of the DNA molecule.

94. A method of claim 93, wherein the region of the DNA molecule that is deleted is prone to cross-over events in meiosis.

95. The method of any one of claims 91-94, wherein the first and second gRNAs of steps (a)-(d) hybridize to the same strand of the DNA, and the third and fourth gRNAs of steps (a)-(d) hybridize to the opposing strand of the DNA.

96. The method of any one of claims 91-95, wherein the gRNAs of steps (a) and (b) hybridize to regions of the DNA that are no more than 100 base pairs apart, and the gRNAs of steps (c) and (d) hybridize to regions of the DNA that are no more than 100 base pairs apart.

97. The method of claim 96, wherein the gRNAs of steps (a) and (b) hybridize to regions of the DNA that are no more than 50, no more than 60, no more than 70, no more than 80, or no more than 90 base pairs apart, and the gRNAs of steps (c) and (d) hybridize to regions of the DNA that are no more than 10, no more than 15, no more than 20, no more than 25, no more than 30, no more than 40, no more than 50, no more than 60, no more than 70, no more than 80, or no more than 90 base pairs apart.

98. The method of any one of claims 91-97, wherein the gRNAs of steps (a) and (b); and/or the gRNAs of steps (c) and (d) hybridize to gRNA binding sites flanking a recombinase site.

99. The method of claim 98, wherein the recombinase site comprises a res, gix, hix, six, resH, LoxP, FTR, or att core or related core sequence.

100. The method of claim 98 or claim 99, wherein the recombinase site comprises a gix core or gix-related core sequence.

101. The method of claim 99 or claim 100, wherein the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 3 to 7 base pairs.

102. The method of any one of claims 99-101, wherein the distance between the gix core or gix-related core sequence and at least one gRNA binding site is from 5 to 6 base pairs.

103. The method of any one of claims 81-102, wherein the DNA is in a cell.

104. The method of claim 103, wherein the cell is a eukaryotic cell.

105. The method of claim 103, wherein the cell is a prokaryotic cell.

106. The method of claim 104 or claim 105, wherein the cell is in a subject.

107. The method of claim 106, wherein the subject is a human.

108. A polynucleotide encoding a fusion protein of any one of claims 1-63.

109. A vector comprising a polynucleotide of claim 108.

110. A vector for recombinant protein expression comprising a polynucleotide encoding a fusion protein of any one of claims 1-63.

111. A cell comprising a genetic construct for expressing a fusion protein of any one of claims 1-63.

112. A kit comprising a fusion protein of any one of claims 1-63.

113. A kit comprising a polynucleotide encoding a fusion protein of any one of claims 1-63.

114. A kit comprising a vector for recombinant protein expression, wherein the vector comprises a polynucleotide encoding a fusion protein of any one of claims 1-63.

115. A kit comprising a cell that comprises a genetic construct for expressing a fusion protein of any one of claims 1-63.

116. The kit of any one of claims 112-115, further comprising one or more gRNAs and/or vectors for expressing one or more gRNAs.

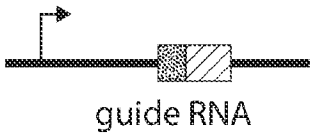


Figure 1A

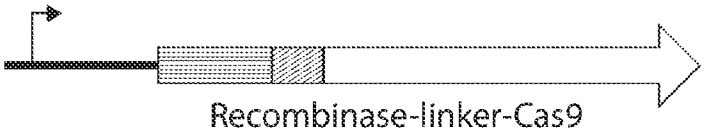


Figure 1B

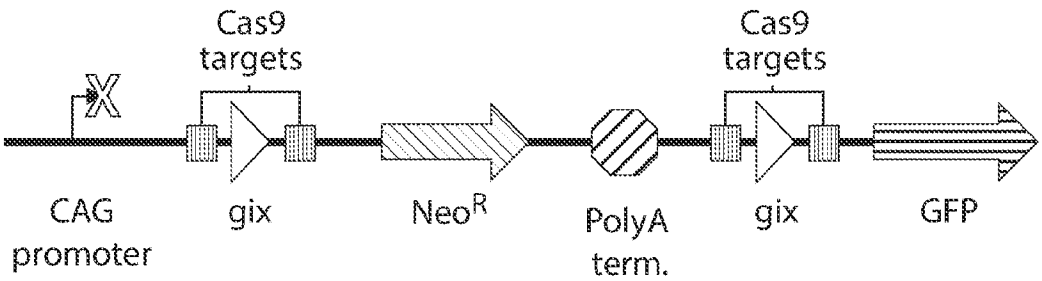


Figure 1C

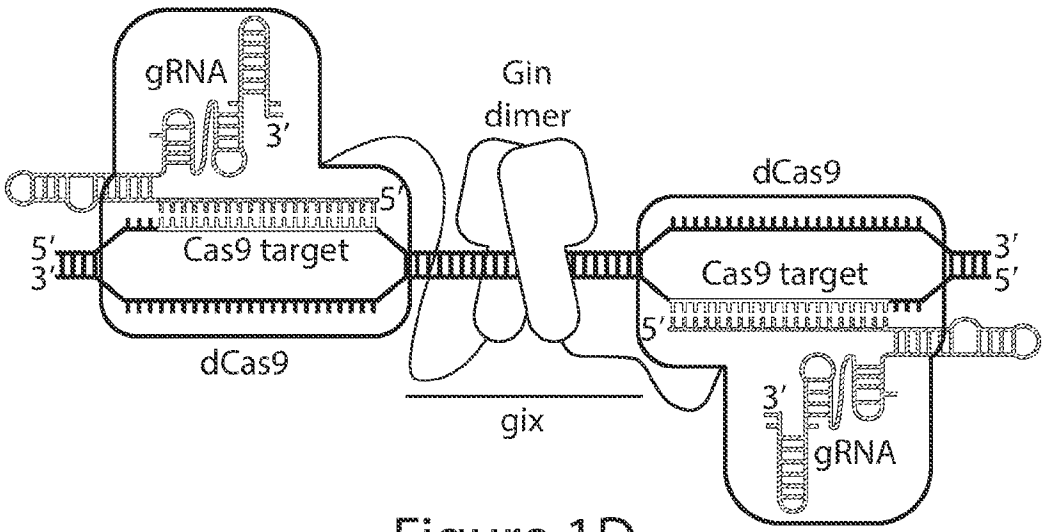


Figure 1D

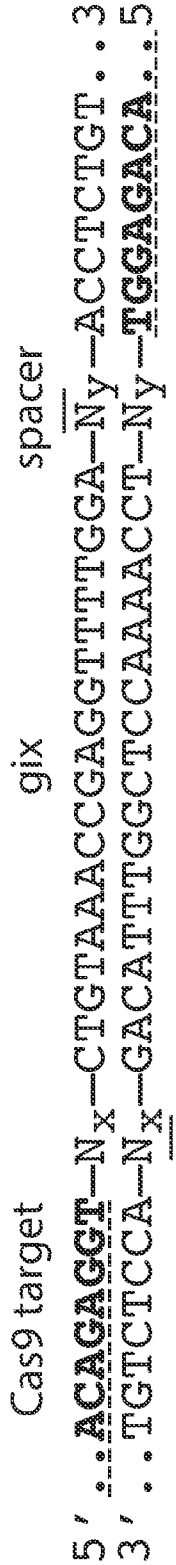


Figure 2A

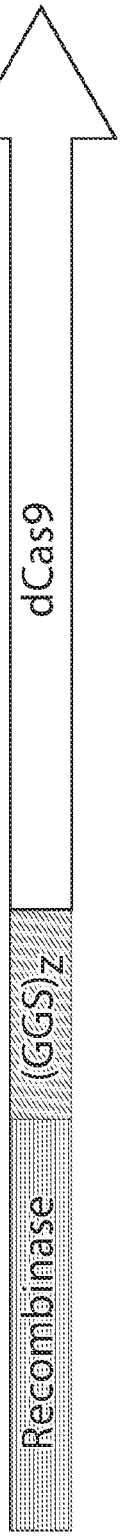


Figure 2B

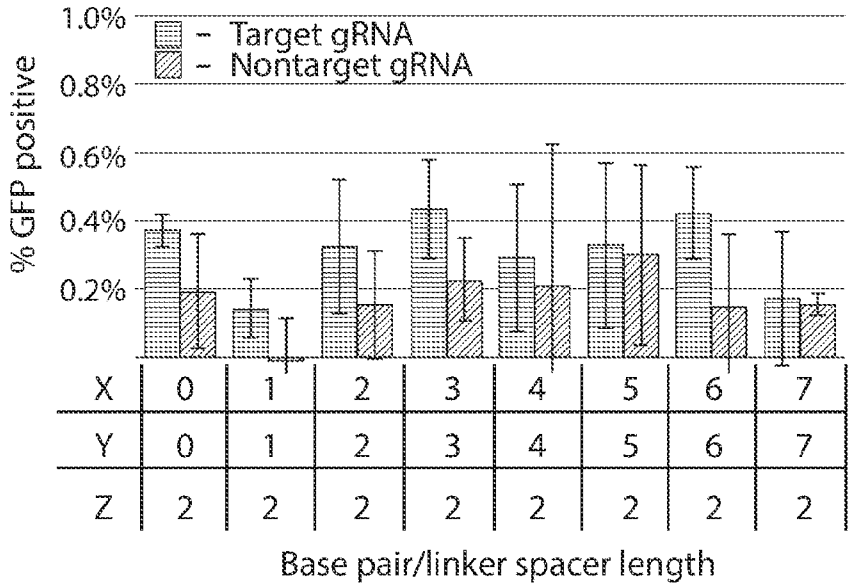


Figure 2C

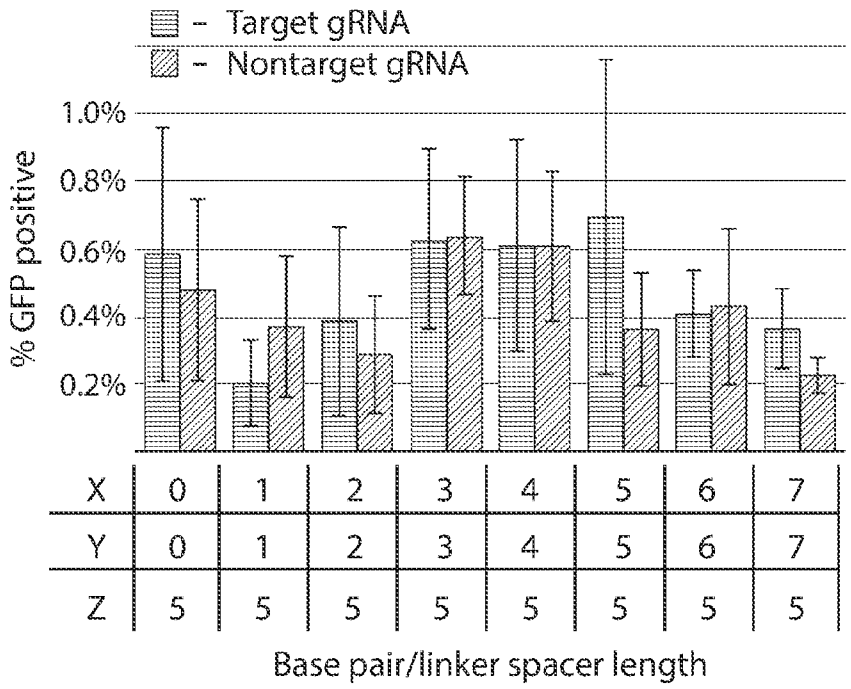


Figure 2D

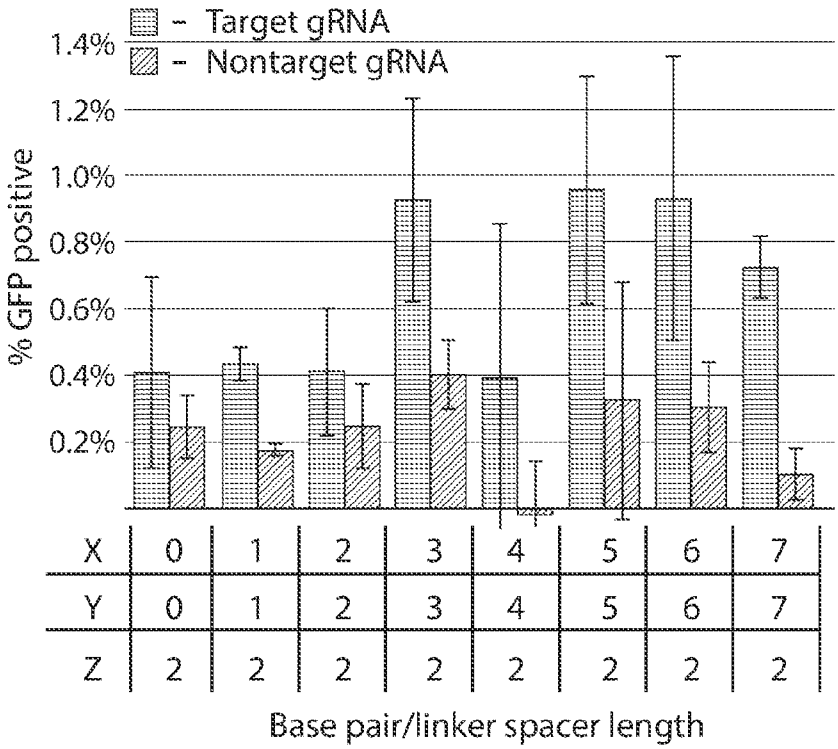


Figure 2E

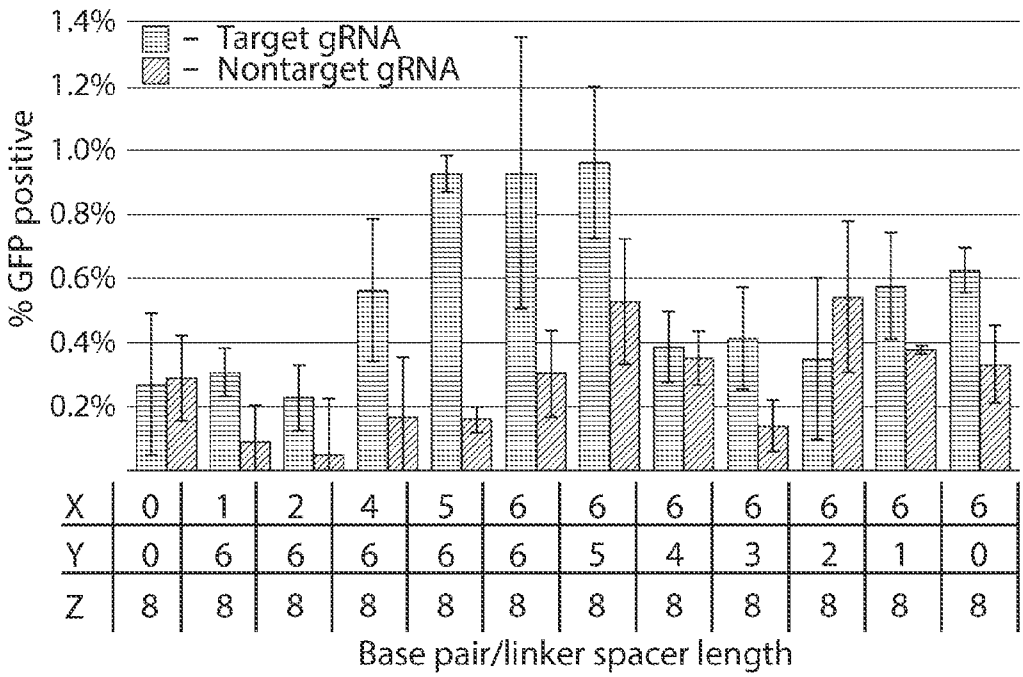


Figure 2F

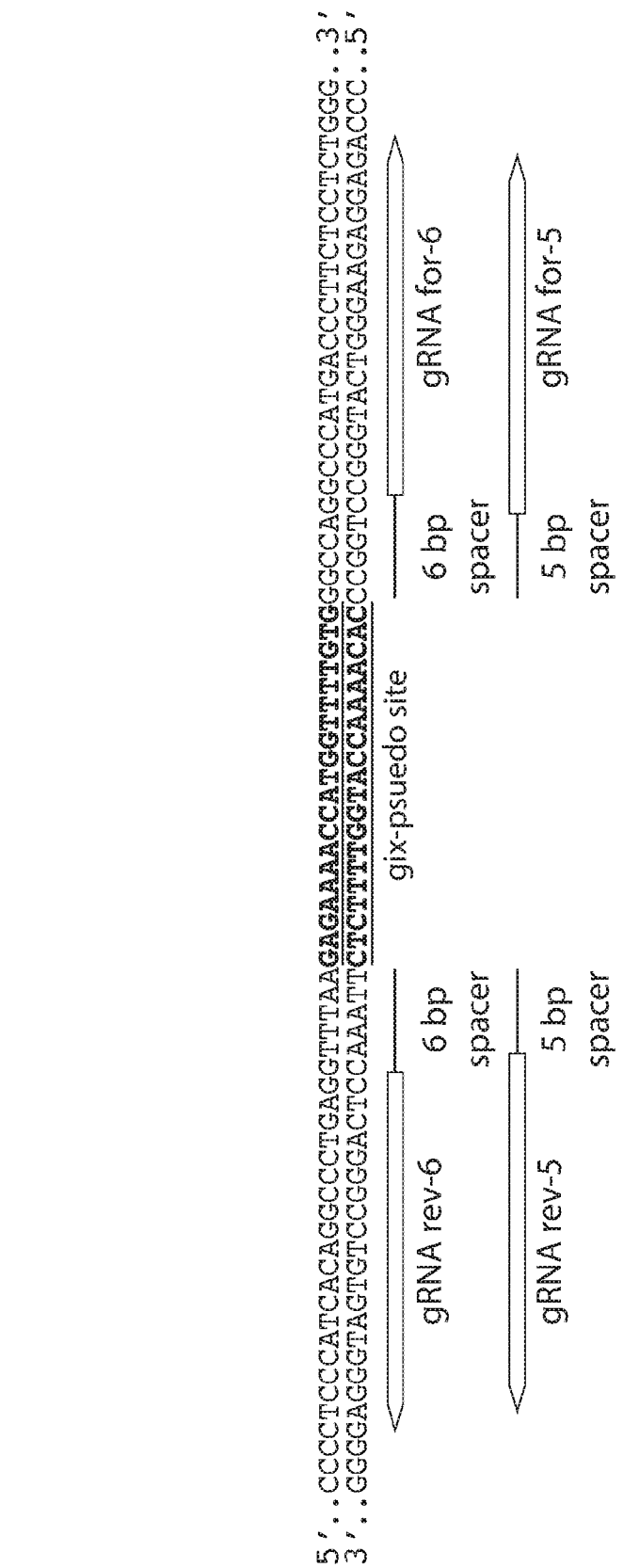

5'..CCCCTCCCATCACAGGCCCTGAGGTTTAAGAGAAACCATGGTTTTGTGGGCCAGGCCCATGACCCTTCTCCTCTGGG..3'
3'..GGGGAGGGTAGTGTCCGGGACTCCAAATTCTCTTTTGGTACCAAAACACCCGGTCCGGGTACTGGGAAGAGGAGACCC..5'
gix-psuedo site
gRNA rev-6
6 bp spacer
gRNA for-6
6 bp spacer
gRNA rev-5
5 bp spacer
gRNA for-5
5 bp spacer

Figure 3A

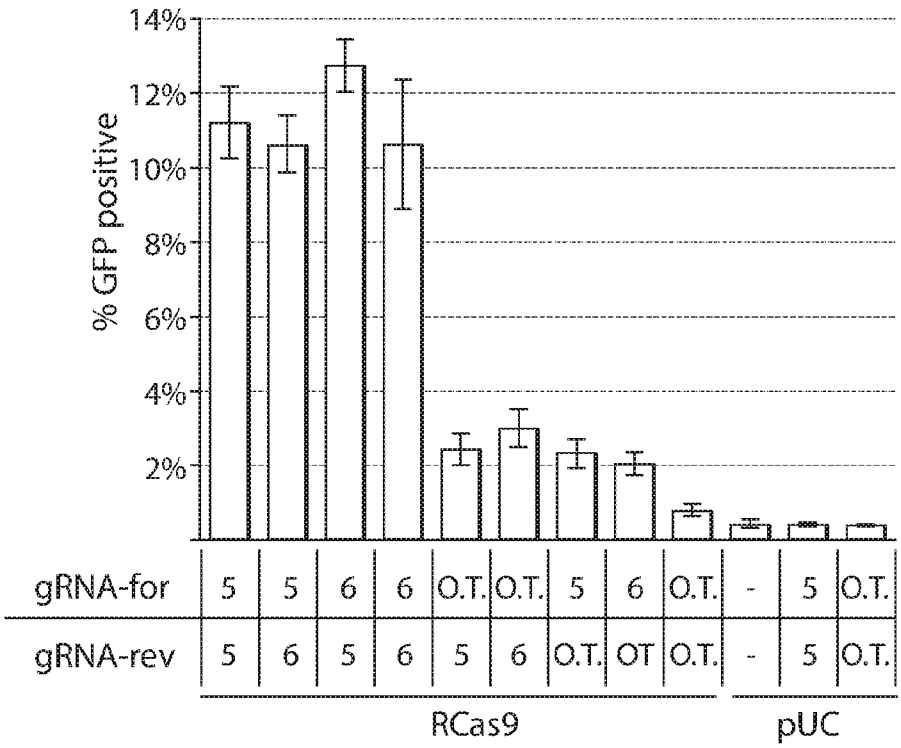

14%
12%
10%
8%
6%
4%
2%
% GFP positive
gRNA-for 5 5 6 6 O.T. O.T. 5 6 O.T. - 5 O.T.
gRNA-rev 5 6 5 6 5 6 O.T. OT O.T. - 5 O.T.
RCas9
pUC

Figure 3B

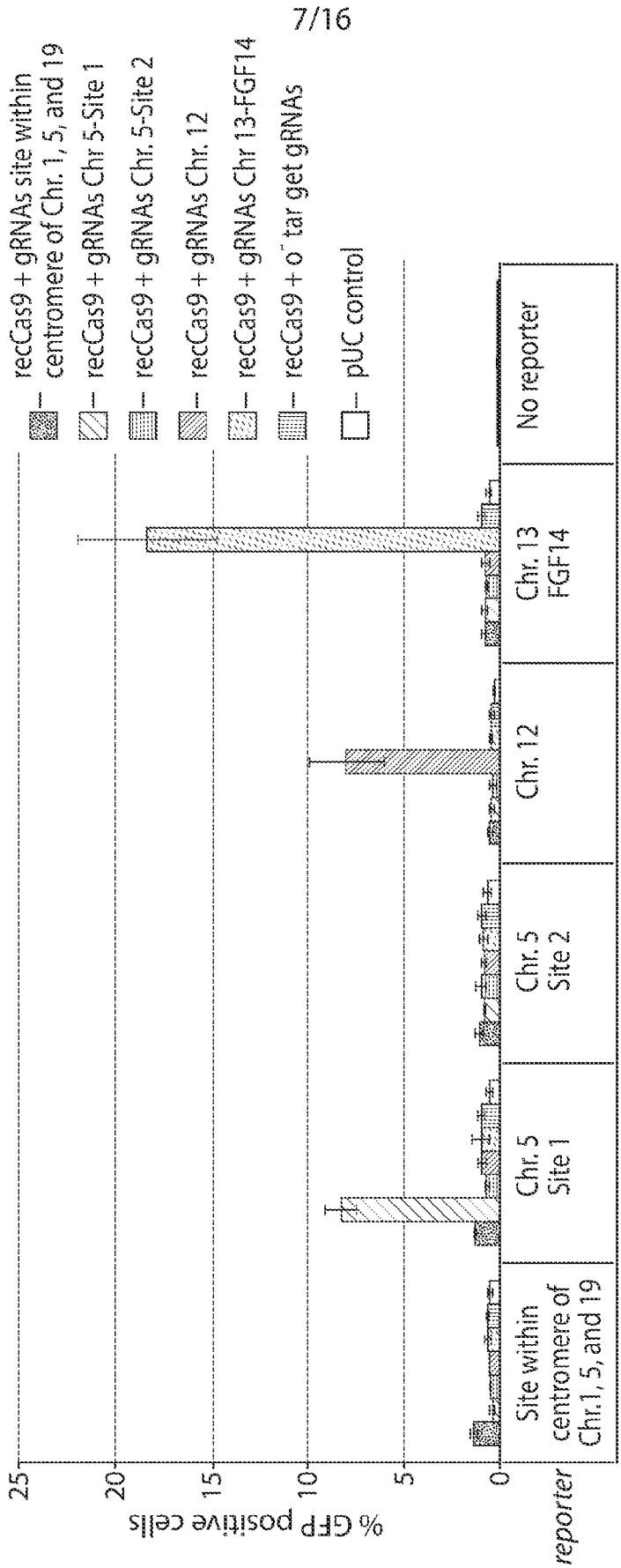
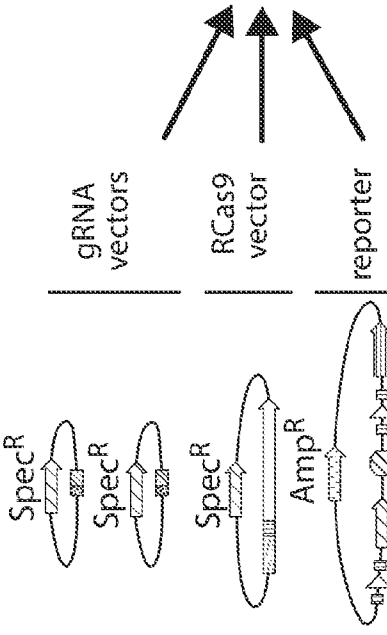

recCas9 + gRNAs site within centromere of Chr. 1, 5, and 19
recCas9 + gRNAs Chr 5-Site 1
recCas9 + gRNAs Chr. 5-Site 2
recCas9 + gRNAs Chr. 12
recCas9 + gRNAs Chr 13-FGF14
recCas9 + o⁻ tar get gRNAs
pUC control
% GFP positive cells
25
20
15
10
5
0
reporter
Site within centromere of Chr.1, 5, and 19
Chr. 5 Site 1
Chr. 5 Site 2
Chr. 12
Chr. 13 FGF14
No reporter

Figure 4A



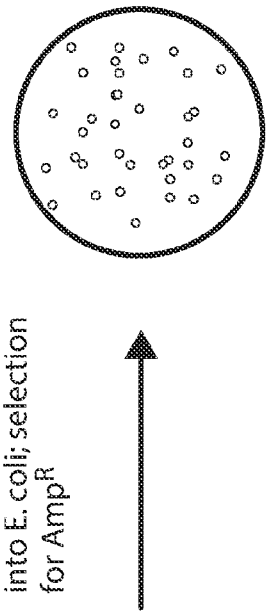


Figure 4B

gRNA/RCas9 Combinations

Reporter		gRNAs Chr. 5-site 1	gRNAs Chr. 12	gRNAs Chr. 13-FGF14	pUC control
	GinB-8GGS--dCas9				
Reporter: Chr. 5 Site 1		11.96±0.54%	0.00%	0.00%	0.00%
Reporter: Chr. 12		0.00%	23.49±0.41%	0.00%	0.00%
Reporter: Chr. 13 FGF14		0.00%	0.00%	31.73±4.27%	0.00%

Figure 4D

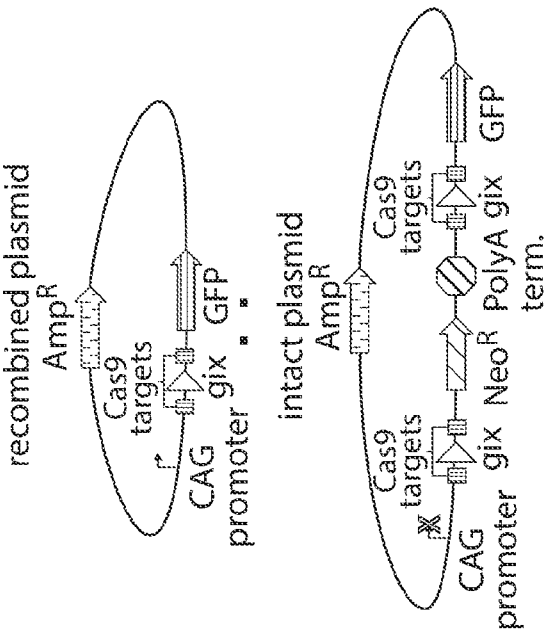


Figure 4C

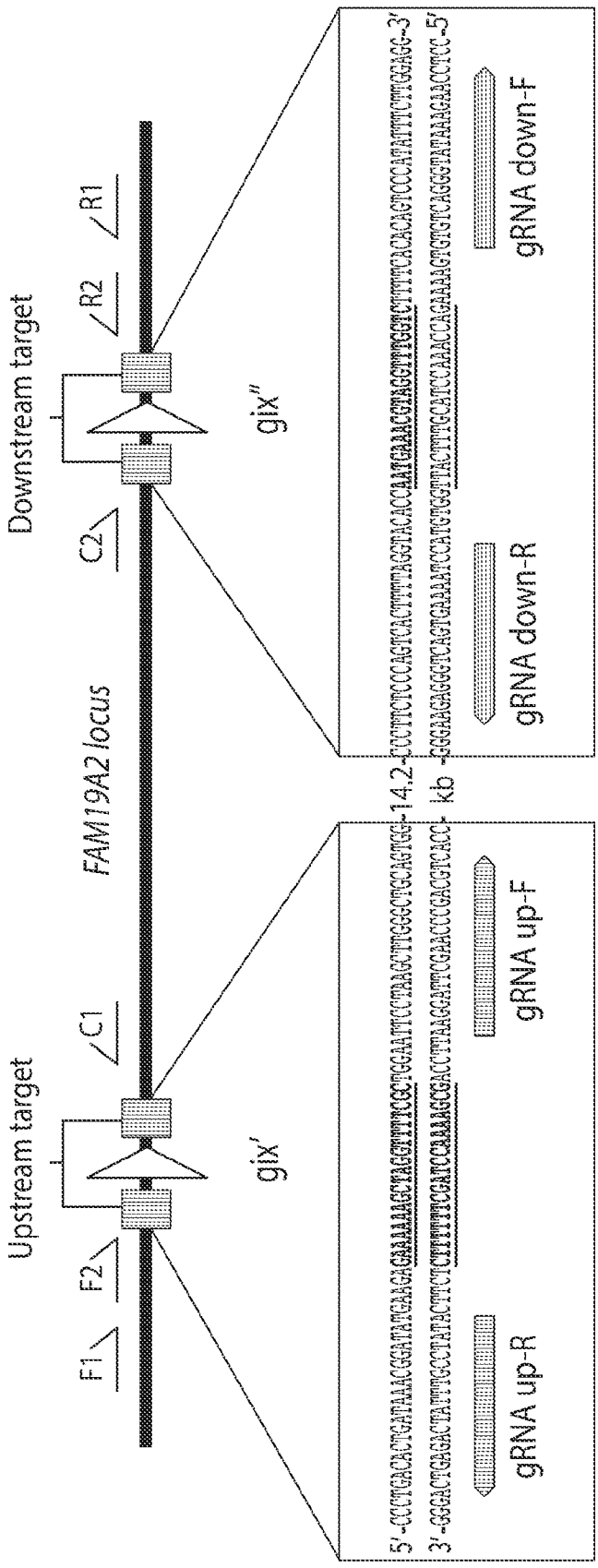

Upstream target
Downstream target
FAM19A2 locus
F1
F2
C1
C2
R1
R2
gix'
gix''
gRNA up-R
gRNA up-F
gRNA down-R
gRNA down-F
14.2 kb

Figure 5A

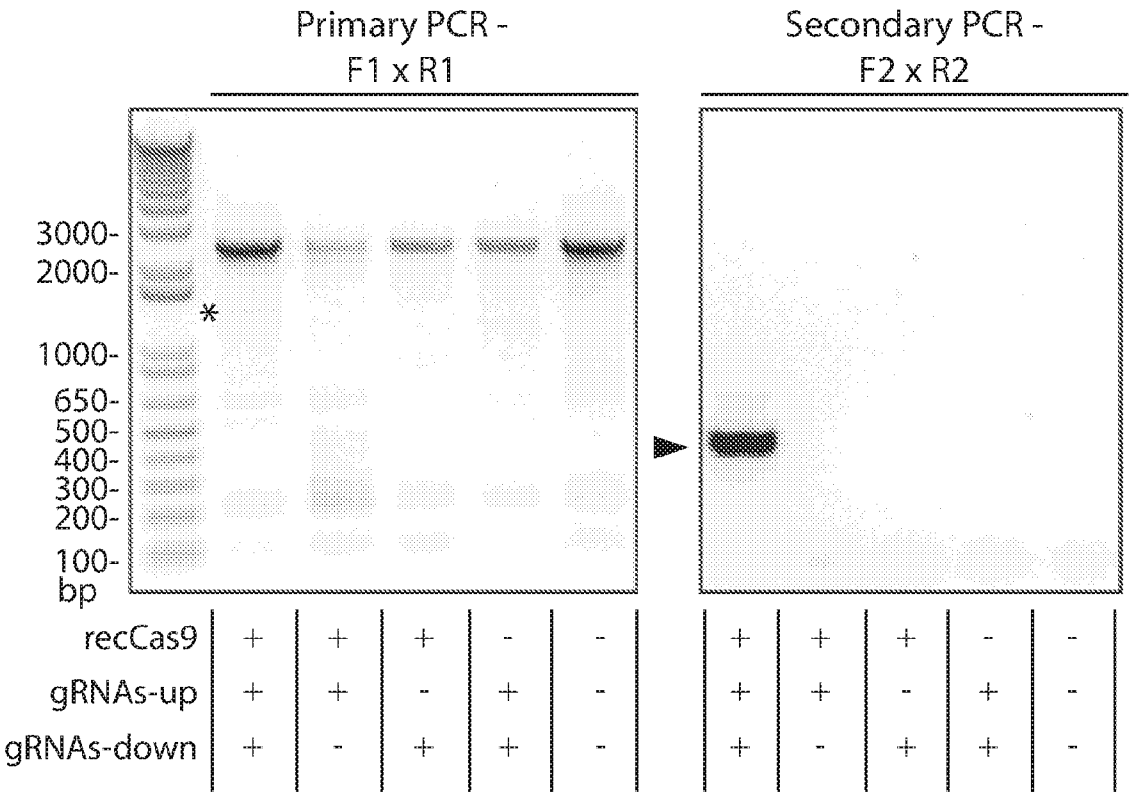

Primary PCR -
F1 x R1
Secondary PCR -
F2 x R2
3000-
2000-
1000-
650-
500-
400-
300-
200-
100-
bp
*
recCas9 + + + - - + + + - -
gRNAs-up + + - + - + + - + -
gRNAs-down + - + + - + - + + -

Figure 5B

Figure 5C

Nested PCR template

	Sample 1	Sample 2	Sample 3	Untransfected control
Minimum Deletion	0.036±0.0233%	0.011±0.0072%	0.021±0.0091%	<0.0072%

Figure 5D

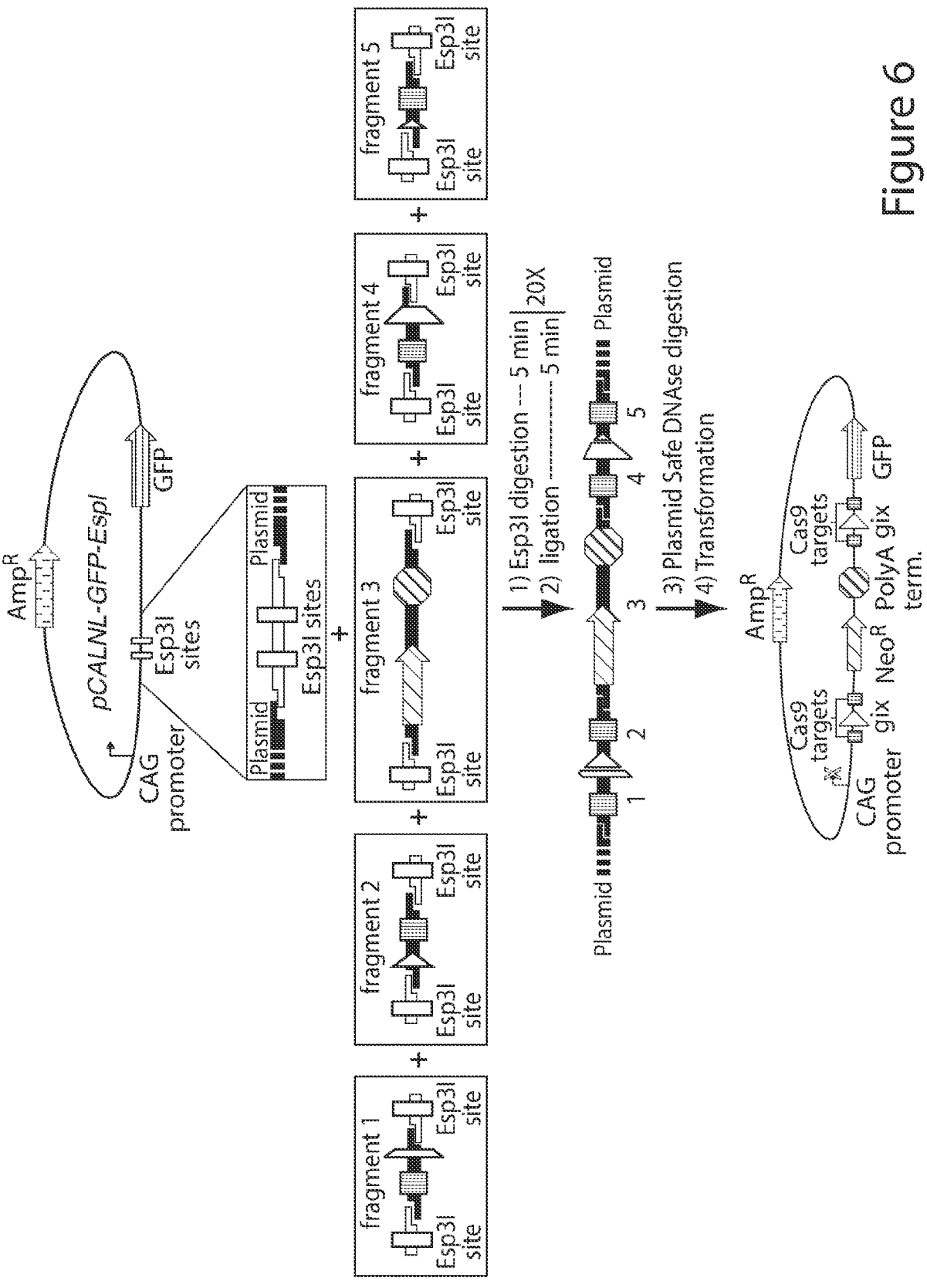

Amp^R
pCALNL-GFP-EspI
GFP
CAG promoter
Esp3I sites
Plasmid
Esp3I sites
Plasmid
fragment 1
fragment 2
fragment 3
fragment 4
fragment 5
Esp3I site
1) Esp3I digestion --- 5 min
2) ligation ---------- 5 min
20X
Plasmid
1
2
3
4
5
3) Plasmid Safe DNAse digestion
4) Transformation
Amp^R
Cas9 targets
CAG promoter
gix
Neo^R
PolyA
gix
Cas9 targets
GFP
term.

Figure 6

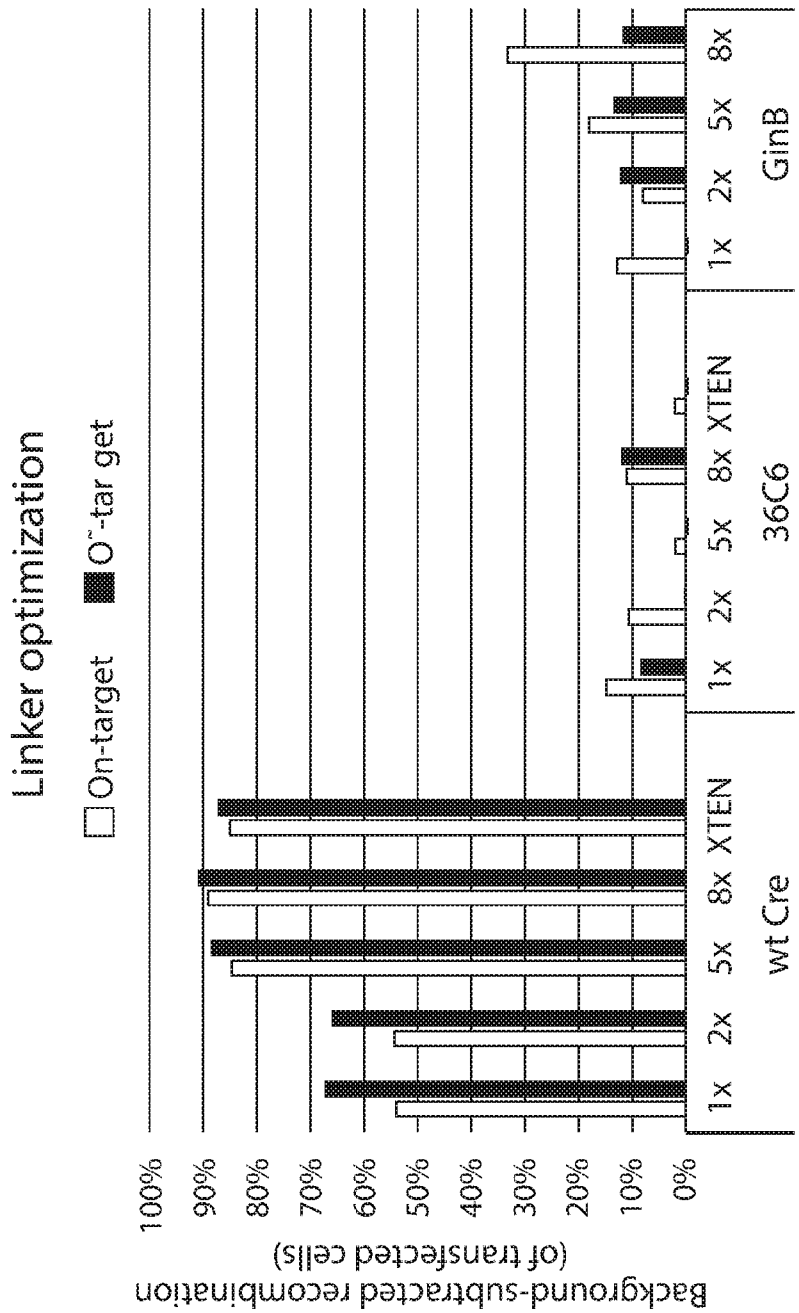

Linker optimization
On-target
O˜-tar get
Background-subtracted recombination
(of transfected cells)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1x 2x 5x 8x XTEN
wt Cre
1x 2x 5x 8x XTEN
36C6
1x 2x 5x 8x
GinB

Figure 7A

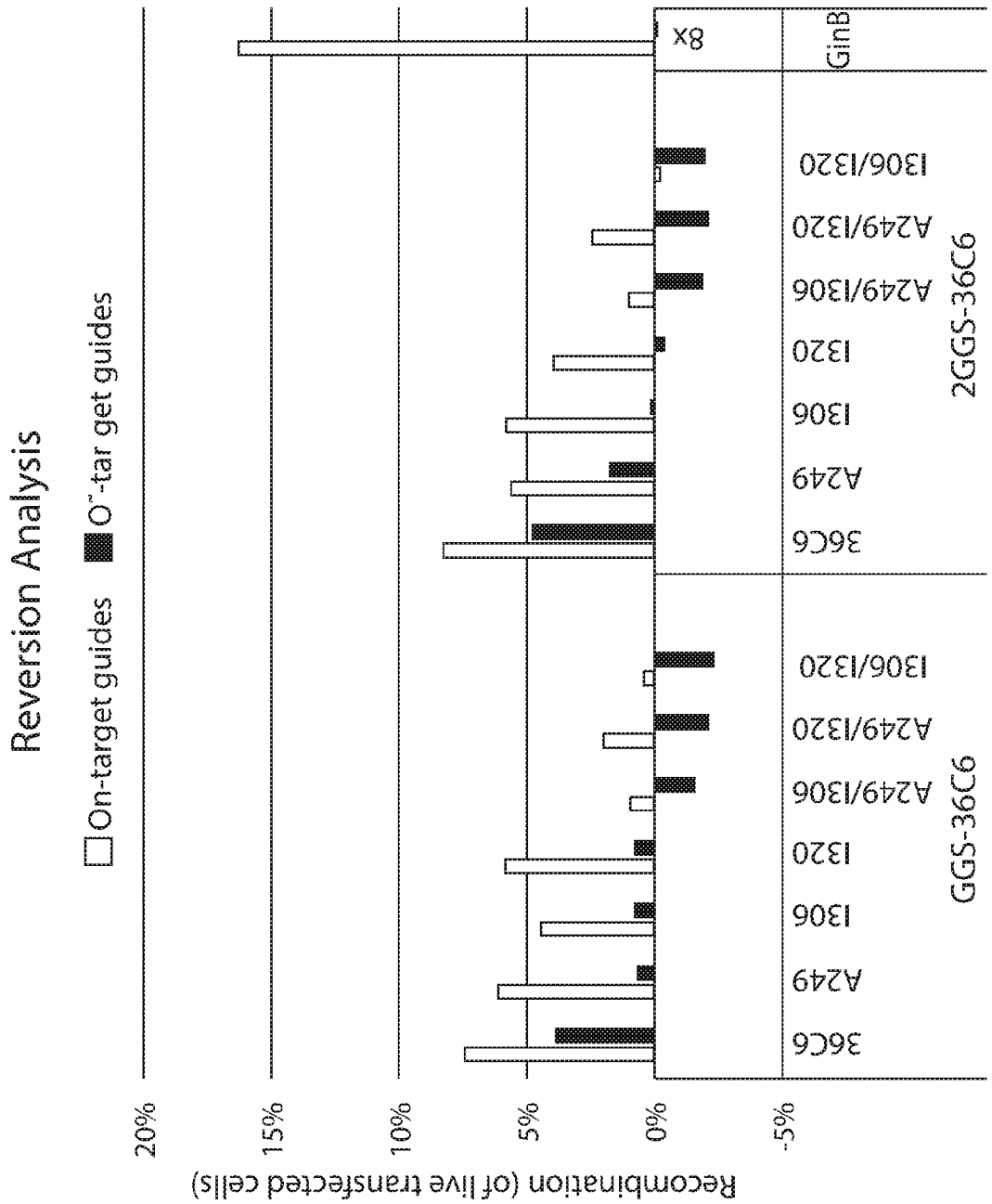
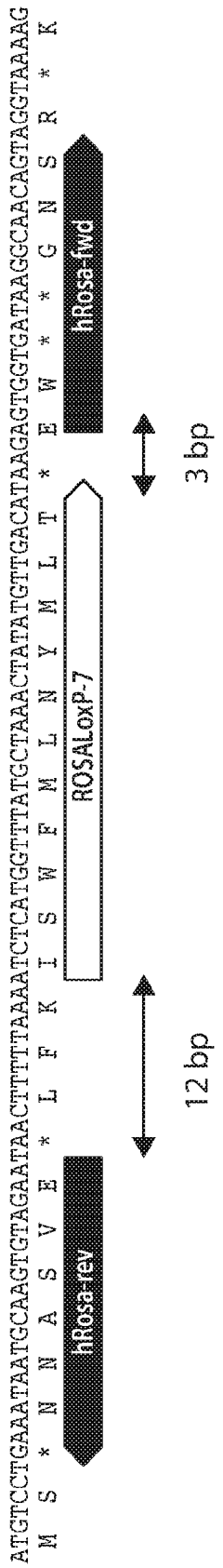

Reversion Analysis
On-target guides
O˜-tar get guides
Recombination (of live transfected cells)
20%
15%
10%
5%
0%
-5%
GinB
8x
2GGS-36C6
GGS-36C6
36C6
A249
I306
I320
A249/I306
A249/I320
I306/I320

Figure 7B


ATGTCCTGAAATAATGCAAGTGTAGAATAACTTTTTAAAATCTCATGGTTTATGCTAAACTATATGTTGACATAAGAGTGGTGATAAGGCAACAGTAGGTAAAAG
M S * N N A S V E * L F K I S W F M L N Y M L T * E W * * G N S R * K
hRosa-rev
ROSALoxP-7
hRosa-fwd
12 bp
3 bp

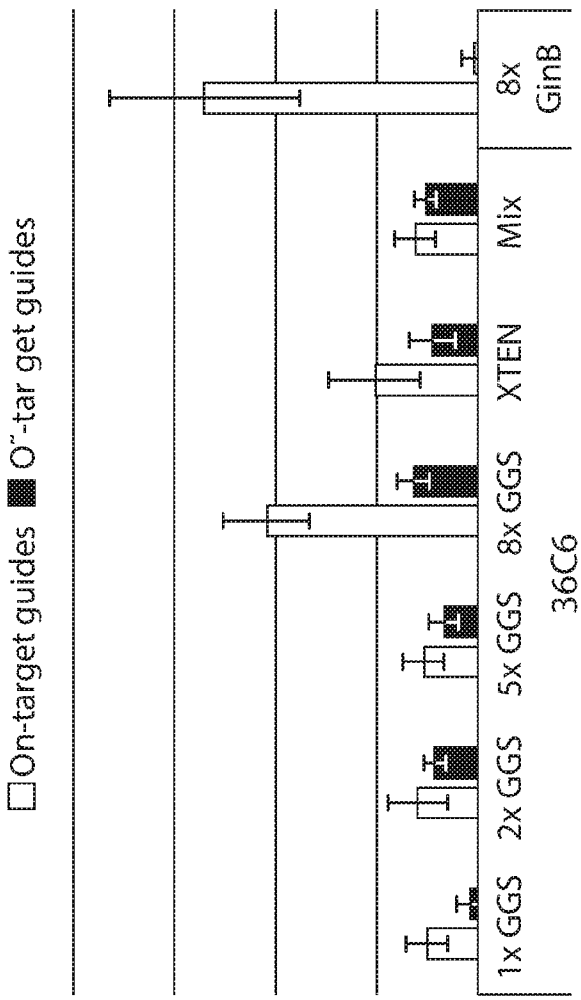

On-target guides
O˜-tar get guides
20%
15%
10%
5%
0%
1x GGS
2x GGS
5x GGS
8x GGS
XTEN
Mix
8x GinB
36C6

Figure 8

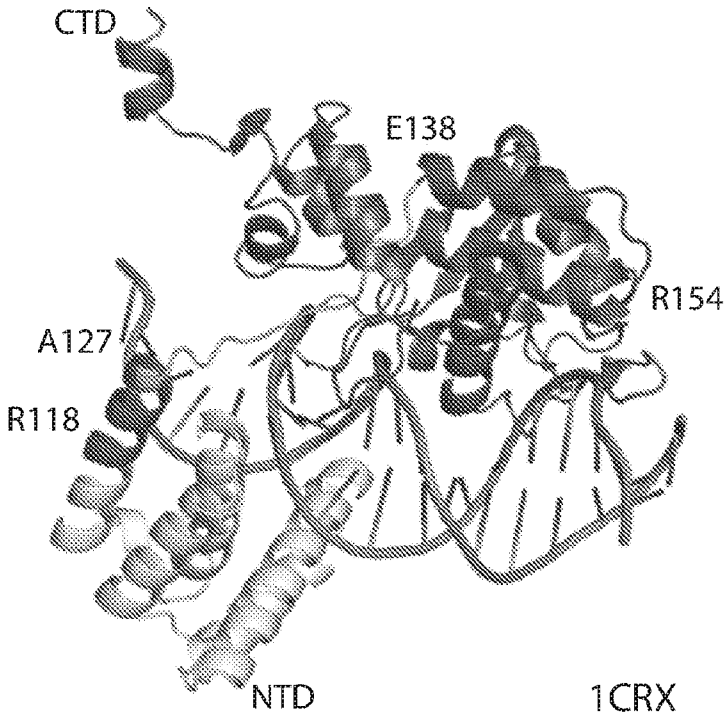


Figure 9A

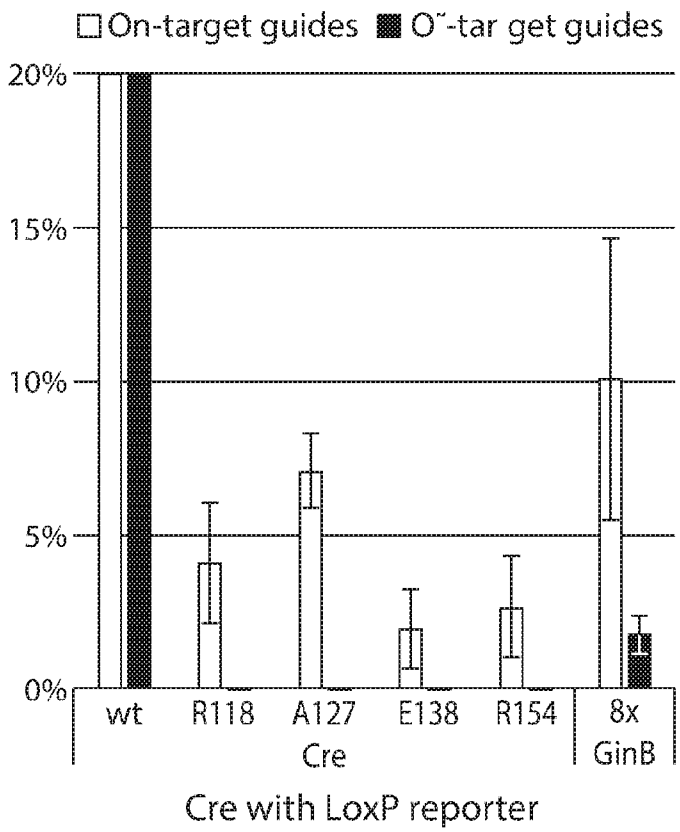


Figure 9B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/046144

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

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

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

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

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

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

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

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

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/046144

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/22 C12N15/11 A61K48/00
ADD.

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)
C12N A61K

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2015/071898 A1 (LIU DAVID R [US] ET AL) 12 March 2015 (2015-03-12) Whole document, especially the claims. -----	1-13, 19-22, 32,64, 66-68, 77, 108-113
A	YUAN P ET AL: "Tetrameric Structure of a Serine Integrase Catalytic Domain", STRUCTURE, ELSEVIER, AMSTERDAM, NL, vol. 16, no. 8, 6 August 2008 (2008-08-06) , pages 1275-1286, XP025884134, ISSN: 0969-2126, DOI: 10.1016/J.STR.2008.04.018 [retrieved on 2008-08-05] the whole document -----	1-116

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 October 2017	10/10/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Kools, Patrick

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GORDLEY RUSSELL M ET AL: "Synthesis of programmable integrases", PROCEEDINGS NATIONAL ACADEMY OF SCIENCES PNAS, NATIONAL ACADEMY OF SCIENCES, US, vol. 106, no. 13, 1 March 2009 (2009-03-01), pages 5053-5058, XP002544501, ISSN: 0027-8424, DOI: 10.1073/PNAS.0812502106 the whole document -----	1-116
A	YUEJU WANG ET AL: "Recombinase technology: applications and possibilities", PLANT CELL REPORTS, SPRINGER, BERLIN, DE, vol. 30, no. 3, 24 October 2010 (2010-10-24), pages 267-285, XP019880902, ISSN: 1432-203X, DOI: 10.1007/S00299-010-0938-1 the whole document -----	1-116
A	T. GAJ ET AL: "Structure-guided reprogramming of serine recombinase DNA sequence specificity", PROCEEDINGS NATIONAL ACADEMY OF SCIENCES PNAS, vol. 108, no. 2, 27 December 2010 (2010-12-27), pages 498-503, XP055411390, US ISSN: 0027-8424, DOI: 10.1073/pnas.1014214108 the whole document -----	1-116
X,P	BRIAN CHAIKIND ET AL: "A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells", NUCLEIC ACIDS RESEARCH, vol. 44, no. 20, 11 August 2016 (2016-08-11), pages 9758-9770, XP055411362, ISSN: 0305-1048, DOI: 10.1093/nar/gkw707 the whole document -----	1-116

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2017/046144

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2015071898 A1	12-03-2015	AU 2014315070 A1	17-03-2016
		CA 2923418 A1	12-03-2015
		EP 3041497 A2	13-07-2016
		JP 2016537008 A	01-12-2016
		KR 20160050069 A	10-05-2016
		US 2015071898 A1	12-03-2015
		US 2015071899 A1	12-03-2015
		US 2016215276 A1	28-07-2016
		WO 2015035162 A2	12-03-2015