

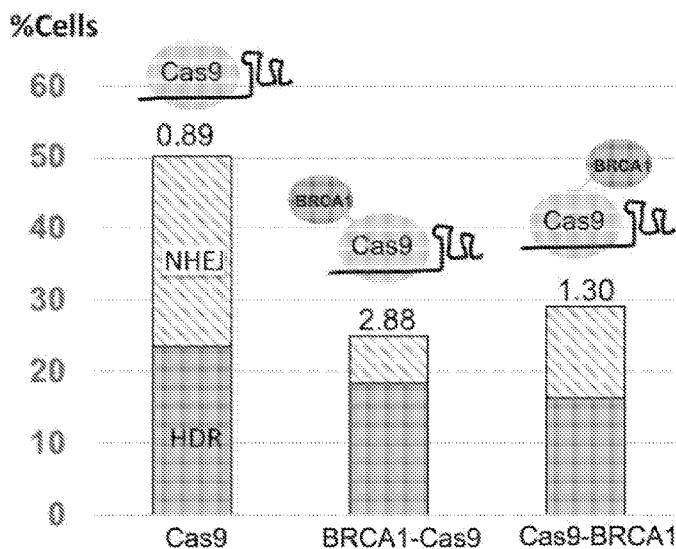


- (51) International Patent Classification:
C12N 9/22 (2006.01) C12N 15/90 (2006.01)
C12N 9/96 (2006.01)
- (21) International Application Number:
PCT/US2019/047021
- (22) International Filing Date:
19 August 2019 (19.08.2019)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
62/720,847 21 August 2018 (21.08.2018) US
- (71) Applicant: **THE JACKSON LABORATORY** [US/US];
600 Main Street, Bar Harbor, ME 04609 (US).
- (72) Inventors: **CHENG, Albert**; 600 Main Street, Bar Harbor,
ME 04609 (US). **JILLETTE, Nathaniel**; 600 Main Street,
Bar Harbor, ME 04609 (US).

- (74) Agent: **DIPIETRANTONIO, Heather, J.** et al.; Wolf,
Greenfield & Sacks, P.C., 600 Atlantic 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,

(54) Title: METHODS AND COMPOSITIONS FOR RECRUITING DNA REPAIR PROTEINS

FIG. 3



(57) Abstract: The DNA repair complexes provided herein are, *inter alia*, useful for editing genome sequences by introducing precise changes in a target site in the presence of a donor sequence. The RNA-guided DNA endonuclease provided herein including embodiments thereof (e.g., Cas9 nuclease or Cas9 nickase) is capable of introducing a strand break (double- or single-strand break) at a target site in the genome of a cell (e.g., gene or transcriptional regulatory sequence) and the break is then predominantly repaired through the mechanism of HDR. By increasing the HDR efficiency significantly and in certain instances decreasing NHEJ at a target site, the compositions and methods provided herein meet the long-felt need of site directed, highly accurate genome editing.



EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

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

METHODS AND COMPOSITIONS FOR RECRUITING DNA REPAIR PROTEINS**RELATED APPLICATIONS**

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. provisional application number 62/720,847, filed August 21, 2018, which is incorporated by reference herein in its entirety.

BACKGROUND

The development of programmable nucleases has opened an avenue for editing genome sequences of virtually any cell or organism, transforming basic and medical research, and enabling applications from agriculture to therapeutics. Nonetheless, realizing the full potential of this revolutionary technology in patients will depend on highly efficient and specific editing of the genome with nucleotide precision and in sufficient numbers of cells to have therapeutic value. Programmable nucleases introduce a double stranded break (DSB) at a desired site in the genome whereupon the DSB may be repaired by one of a variety of cellular DNA repair mechanisms. The most efficient repair pathway is error-prone non-homologous end joining (NHEJ), which has been widely exploited to disrupt gene function through the introduction of random insertions/deletions (indels).

To introduce precise changes that could be used to correct disease-causing variants, exogenous DNA templates with homology to the target region are required to insert new DNA sequences at the site of the DSB. The cellular machinery used for homology-directed repair (HDR) depends on the repair template itself. While double-stranded supercoiled donor plasmids can undergo classical homologous recombination (HR) with the genome, the mechanism of repair mediated by single-stranded DNA templates is not completely understood and several pathways have been proposed.

NHEJ and HDR represent competing pathways that contain overlapping yet distinct protein components. NHEJ is the dominant outcome even in the presence of a repair template. Attempts to block NHEJ genetically or with protein and chemical inhibitors shift the balance in favor of HDR, but may risk unwanted genome instability. The alternative approach is to stimulate the HDR pathway. Indeed, modest improvement in HDR efficiency has come from the overexpression/activation of proteins specific to the HDR pathway such as RAD51¹⁶ or the fusion of DNA repair proteins with Cas9 nuclease. Other orthogonal strategies that have been reported to slightly improve HDR frequencies include the delivery of recombinant Cas9 ribonucleoprotein, cell cycle control of Cas9 expression or perturbing the cell cycle, cold shock

and tethering the repair template to the nuclease. Despite these advances, NHEJ remains the predominant repair event. Thus, with the added challenge of delivering both the nuclease and the repair template to cells *in vivo*, there is a need in the art to improve HDR efficiency such that it is the preferred repair outcome in treated cells. The compositions and methods provided herein, in some embodiments, address these and other needs in the art.

SUMMARY

Some aspects of the present disclosure provide methods that comprise delivering to a cell comprising a target nucleic acid (a) a ribonucleic acid (RNA)-guided nuclease or a nucleic acid encoding a RNA-guided nuclease, (b) a guide RNA (gRNA) or a nucleic acid encoding a gRNA that comprises (i) a DNA-targeting sequence that can bind specifically to a target nucleic acid sequence, (ii) a RNA-guided nuclease-binding sequence, and (iii) a Pumilio-FBF (PUF) domain binding sequence (PBS), and (c) a DNA repair protein or a nucleic acid encoding a DNA repair protein (e.g., conjugate) that comprises a DNA repair domain linked to a PUF domain that binds to the PBS.

In some embodiments, a target nucleic acid sequence is in the genome of a cell. Thus, a target nucleic acid may be referred to as a genomic target nucleic acid. In some embodiments, a target nucleic acid sequence is within a gene and/or within a transcriptional regulatory sequence.

In some embodiments, the methods comprise delivering to the cell a ribonucleoprotein complex comprising the RNA-guided nuclease (e.g., a Cas9 nuclease or a Cas9 nickase) bound to the gRNA.

In some embodiments, the methods further comprise delivering to the cell a donor nucleic acid comprising a sequence of interest. A donor nucleic acid may be single-stranded or double-stranded, for example.

In some embodiments, the methods further comprise maintaining a cell under conditions that result in cleavage of a target nucleic acid sequence. In some embodiments, the methods further comprise maintaining a cell under conditions that result in the production of a cellular nucleic acid comprising a sequence of interest.

In some embodiments, a RNA-guided DNA nuclease is a Cas9 nuclease or a Cas9 nickase.

In some embodiments, a gRNA comprises at least two PUF domain binding sequences. For example, a gRNA may comprise 2-50 PUF domain binding sequences. In some embodiments, at least two of the PUF binding domain sequences are tandem repeat sequences. In

some embodiments, a PBS has a length of at least 8 nucleotides. In some embodiments, a PBS comprises the nucleotide sequence of SEQ ID NO: 83.

In some embodiments, a PUF domain comprises a PUFa domain, a PUFb domain, a PUFc domain, or a PUFw domain. In some embodiments, a PUF domain comprises a PUFa domain that comprises the amino acid sequence of SEQ ID NO: 27. In some embodiments, a PUF domain comprises a C-terminus and a N-terminus, and the DNA repair domain is linked to the C-terminus or the N-terminus of the PUF domain.

In some embodiments, a DNA repair domain comprises an enzymatic activity selected from the group consisting of ligase activity, polymerase activity, topoisomerase activity, helicase activity, and nuclease activity. In some embodiments, a DNA repair domain comprises a ligase, a polymerase, a topoisomerase, a helicase, or a nuclease. For example, a DNA repair domain may comprise a protein selected from the group consisting of: Replication Protein A1 (RPA1); Replication Protein A2 (RPA2); Fanconi Anemia Complementation Group M (FANCM); RAD51 Recombinase (RAD51); RAD52 Homolog, DNA Repair Protein (RAD52); RAD51 Paralog C (RAD51C); RAD18 E3 Ubiquitin Protein Ligase (RAD18); RB Binding Protein 8, Nuclease (RBBP8/CTIP); Tumor Protein P53 Binding Protein 1 (TP53BP1); BRCA1 DNA Repair Associated (BRCA1); RAD54-like (RAD54L); Partner and Localizer of BRCA2 (PALB2); X-Ray Repair Cross Complementing 3 (XRCC3); MRE11 Homolog, Double Strand Break Repair Nuclease (MRE11A); Flap Structure-Specific Nuclease 1 (FEN1); RecQ Like Helicase 5 (RECQ5); FA Complementation Group B (FANCB); Ubiquitin Specific Peptidase 1 (USP1); FA Complementation Group F (FANCF); and FA Complementation Group G (FANCG).

In some embodiments, a gRNA, a RNA-guided nuclease, a DNA repair protein, and/or a donor nucleic acid are encoded on independent vectors or on the same vector. For example, a vector may be a plasmid or a viral vector.

Other aspects of the present disclosure provide DNA repair proteins (e.g., conjugates) comprising a DNA repair domain linked (directly or indirectly) to a Pumilio-FBF (PUF) domain.

Further aspects provide nucleic acids encoding DNA repair proteins (e.g., conjugates) of the present disclosure.

Still other aspects provide expression vectors comprising a promoter operably linked to nucleic acids encoding DNA repair proteins (e.g., conjugates) of the present disclosure.

Some aspects of the present disclosure provide kits comprising DNA repair proteins (e.g., conjugates), nucleic acids, or expression vectors described herein. In some embodiments, the kits

further comprise a RNA-guided nuclease or a nucleic acid encoding a RNA-guided nuclease. In some embodiments, the kits further comprise a gRNA or a nucleic acid encoding a gRNA that comprises (i) a DNA-targeting sequence that is complementary to a target nucleic acid sequence, (ii) a RNA-guided nuclease-binding sequence, and (iii) a PBS, wherein the PUF domain of the DNA repair protein (e.g., conjugate) can bind to the PBS. In some embodiments, the kits further comprise a donor nucleic acid that comprises a sequence of interest.

Other aspects provide cells comprising a DNA repair protein (e.g., conjugate), a nucleic acid, or an expression vector of the present disclosure. In some embodiments, a cell further comprises a RNA-guided nuclease or a nucleic acid encoding a RNA-guided nuclease (e.g., Cas9 nuclease or Cas9 nickase). In some embodiments, a cell further comprises a gRNA or a nucleic acid encoding a gRNA that comprises (i) a DNA-targeting sequence that is complementary to a target nucleic acid sequence, (ii) a RNA-guided nuclease-binding sequence, and (iii) a PBS to which the PUF domain of the DNA repair protein (e.g., conjugate) can bind.

Also provided herein, in some aspects, are methods comprising delivering to a cell a programmable nuclease-based gene editing system that comprises a programmable nuclease linked to a DNA repair domain, wherein the programmable nuclease cleaves a target nucleic acid sequence, and the DNA repair domain is selected from the group consisting of: RPA1; RPA2; FANCM; BRCA1; RAD54L; PALB2; XRCC3; FEN1; RECQ5; FANCB; USP1; FANCF; and FANCG. In some embodiments, the methods further comprise delivering to the cell a donor nucleic acid comprising a sequence of interest.

In some embodiments, a programmable nuclease comprises a RNA-guided nuclease, such as Cas9 nuclease or Cas9 nickase. In some embodiments, a method further comprises delivering to the cell a gRNA or a nucleic acid encoding a gRNA that specifically binds to a target nucleic acid sequence.

In some embodiments, a programmable nuclease comprises a zinc finger nuclease (ZFN). In other embodiments, a programmable nuclease comprises a transcription activator-like effector nuclease (TALEN).

It should be understood that a programmable nuclease may be directly linked to (e.g., fused to) the DNA repair domain or indirectly linked (e.g., via at least one linker molecule) to the DNA repair domain.

Other aspects of the present disclosure provide DNA repair proteins comprising a programmable nuclease linked to DNA repair domain selected from the group consisting of: RPA1; RPA2; FANCM; BRCA1; RAD54L; PALB2; XRCC3; FEN1; RECQ5; FANCB; USP1;

FANCF; and FANCG. In some embodiments, a programmable nuclease comprises a RNA-guided nuclease (e.g., Cas9 nuclease or Cas9 nickase), a ZFN, or a TALEN.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1B. The figures show HDR/NHEJ reporters in HEK293T. (Fig. 1A) A constitutively expressed BFP inserted at the AAVS1 locus serves as a gene editing reporter that can be targeted with CRISPR/Cas9 and a BFP->GFP donor to produce non-fluorescent (NHEJ) or green-fluorescent (HDR) cell. (Fig. 1B) Flow cytometry plots of non-targeted HEK293T/BFP cells, or those targeted with Cas9 + sgRNA + donor.

FIG. 2. The figure shows recruitment of DNA repair protein by direct fusion to Cas9. (i) NHEJ is the predominant choice of repair pathway after Cas9 mediated DSB; (ii) fusion of a particular DNA repair protein (DRP) to Cas9 can bias editing outcome towards HDR.

FIG. 3. The figure shows fusion of BRCA1 to Cas9 biases editing outcome towards HDR but with a decrease in total editing. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Unfused Cas9; (B) BRCA1-Cas9; (C) Cas9-BRCA1. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 4. The figure shows recruitment of DNA repair protein to the CRISPR/Cas9 complex via the Casilio methodology. Fusions of DNA repair proteins with Pumilio/FBF (PUF) RNA binding domains can be recruited to the CRISPR/Cas9 complex via binding to the multiple (N) copies pumilio binding sites (PBS) inserted on the single-guide RNA with PBS sites (sgRNA-PBS). Recruitment of particular DRPs (DNA Repair Proteins) can favor editing outcome towards HDR.

FIG. 5. The figures show recruitment of BRCA1 by Casilio strategy enhances HDR without compromising total editing. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Unfused Cas9; (B) BRCA1-Cas9; (C) Cas9-BRCA1; (D) Cas9+BRCA1-PUFa; and (E) Cas9+PUFa-BRCA1. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 6. The figures show recruitment of RAD54L enhances HDR. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9; (B) Cas9+RAD54L-PUFa. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 7. The figure shows recruitment of multiple DNA repair proteins to the CRISPR/Cas9 complex via the Casilio methodology. A casilio complex with Cas9 enzyme can recruit multiple different protein factors (P1, P2, and Pi) to enhance HDR.

FIG. 8. The figures show recruitment of CtIP(T847E)-PALB2(KR)-BRCA1 complex enhances HDR. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9; (B) Cas9+ CtIP(T847E)-PUFa+PALB2(KR)-PUFa+BRCA1-PUFa and (C) Cas9+PUFa-CtIP(T847E)+PUFa-PALB2(KR)+PUFa-BRCA1. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion). Numbers above column indicate HDR/NHEJ ratios.

FIG. 9. The figures show recruitment of RAD51 enhances HDR at nick mediated by Cas9Nickase (Cas9n; Cas9 D10A nickase). BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9Nickase; (B) Cas9Nickase+RAD51-PUFa; or (C) Cas9Nickase+PUFa-RAD51. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 10. The figure shows local recruitment of CtIP(T847E)-PALB2(KR)-BRCA1 complex enhances HDR at nick mediated by Cas9Nickase (Cas9n; Cas9 D10A nickase). BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9Nickase; (B) Cas9Nickase+CtIP(T847E)-PUFa+PALB2(KR)-PUFa+BRCA1-PUFa; or (C) Cas9Nickase+PUFa-CtIP(T847E)+PUFa-PALB2(KR)+PUFa-BRCA1. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 11. The figure shows recruitment of XRCC3 by Casilio strategy enhances HDR. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9; (B) Cas9+XRCC3-PUFa; and (C) Cas9+PUFa-XRCC3. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 12. The figure shows recruitment of RECQ5 by Casilio strategy enhances HDR. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9; (B) Cas9+RECQ5-PUFa; and (C) Cas9+PUFa-RECQ5. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 13. The figure shows recruitment of FEN1 by Casilio strategy enhances HDR. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9; (B) Cas9+FEN1-PUFa; and (C) Cas9+PUFa-FEN1. Stacked columns show the percentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 14. The figure shows recruitment of Fanconi Anemia (FA) pathway proteins by Casilio strategy enhances HDR. BFP->GFP editing experiments were conducted on the HEK293T/BFP reporter cell line using different CRISPR/Cas9 complexes formed by a BFP->GFP ssODN donor, BFP-targeting sgRNA-5xPBSa plus one of (A) Cas9; (B) Cas9+PUFa-FANCB; (C) Cas9+FANCF-PUFa; (D) Cas9+PUFa-FANCF; (E) Cas9+FANCG-PUFa; (F) Cas9+PUFa-FANCG; (G) Cas9+FANCM-PUFa; and (H) Cas9+PUFa-FANCM. Stacked columns show the polynucpercentage of GFP+ cells, indicative of HDR (green lower portion) and percentage of double negative (Dbl-) cells, indicative of NHEJ (patterned upper portion).

FIG. 15. The figure shows more examples of factors that enhance HDR when recruited to site of Cas9Nickase (Cas9n; Cas9 D10A nickase)-mediated DNA nick.

DETAILED DESCRIPTION

Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art. See, e.g., Singleton et al., *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY* 2nd ed., J. Wiley &

Sons (New York, NY 1994); Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, Cold Springs Harbor Press (Cold Springs Harbor, NY 1989). Any methods, devices and materials similar or equivalent to those described herein can be used as provided herein. The following definitions are provided to facilitate understanding of certain terms used frequently herein and are not meant to limit the scope of the present disclosure.

The terms “nucleic acid” and “polynucleotide” may be used interchangeably herein.

Nucleic acids, including nucleic acids with a phosphothioate backbone can include one or more reactive moieties. As used herein, the term reactive moiety includes any group capable of reacting with another molecule, e.g., a nucleic acid or polypeptide through covalent, non-covalent or other interactions. By way of example, the nucleic acid can include an amino acid reactive moiety that reacts with an amino acid on a protein or polypeptide through a covalent, non-covalent or other interaction.

Nucleic acids may include nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphodiester derivatives including, e.g., phosphoramidate, phosphorodiamidate, phosphorothioate (also known as phosphothioate), phosphorodithioate, phosphonocarboxylic acids, phosphonocarboxylates, phosphonoacetic acid, phosphonoformic acid, methyl phosphonate, boron phosphonate, or O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press); and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, modified sugars, and non-ribose backbones (e.g. phosphorodiamidate morpholino oligos or locked nucleic acids (LNA)), including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, *Carbohydrate Modifications in Antisense Research*, Sanghui & Cook, eds. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, e.g., to increase the stability and half-life of such molecules in physiological environments or as probes on a biochip. Mixtures of naturally occurring nucleic acids and analogs can be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made. In some embodiments, the internucleotide linkages in DNA are phosphodiester, phosphodiester derivatives, or a combination of both.

As used herein, the range of values provided includes the specified value. As recognized by a person of ordinary skill in the art such specified value would reasonably include a standard deviation using measurements generally acceptable in the art. In some embodiments, the standard deviation includes a range extending to +/- 10% of the specified value.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues, wherein the polymer may be linked to (e.g., conjugated to) a moiety that does not include amino acids. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. The terms apply to macrocyclic peptides, peptides that have been modified with non-peptide functionality, peptidomimetics, polyamides, and macrolactams. A protein conjugate may include two or more protein domains directly or indirectly linked to each other. A fusion protein is an example of a protein conjugate. A "fusion protein" refers to a chimeric protein encoding two or more separate protein sequences (e.g., domains) that are recombinantly expressed as a single moiety.

The term "peptidyl" and "peptidyl moiety" means a monovalent peptide.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid. The terms "non-naturally occurring amino acid" and "unnatural amino acid" refer to amino acid analogs, synthetic amino acids, and amino acid mimetics which are not found in nature.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical

Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

An amino acid or nucleotide base "position" is denoted by a number that sequentially identifies each amino acid (or nucleotide base) in the reference sequence based on its position relative to the N-terminus (or 5'-end). Due to deletions, insertions, truncations, fusions, and the like that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence determined by simply counting from the N-terminus will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where a variant has a deletion relative to an aligned reference sequence, there will be no amino acid in the variant that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to a numbered amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

The term "corresponding to," when used in the context of the numbering of a given amino acid or nucleic acid sequence, refers to the numbering of the residues of a specified reference sequence when the given amino acid or nucleic acid sequence is compared to the reference sequence.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent

variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual probe sequences.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the present disclosure.

The following eight groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and

In accordance with its plain ordinary meaning within Chemistry and Biology the term "functional fragment" refers to a protein, peptide, peptidyl moiety or nucleic acid that is comparable in function to another protein, peptide or nucleic acid (i.e., a so-called "reference" protein, peptide or nucleic acid), but differs in composition (e.g., has a similar but not identical amino acid sequence, nucleotide sequence or lengths thereof) and differs in structure or origin to a reference protein, peptide or nucleic acid. Thus, the term "functional fragment" includes any recombinant or naturally-occurring form of a protein or nucleic acid sequence, variants thereof that maintain protein or nucleic acid sequence activity (e.g. within at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% activity compared to wildtype protein or nucleic acid sequence). One of skill will recognize that a functional fragment of a protein or nucleic acid may include individual substitutions, deletions or additions to a protein or nucleic acid, which alters, adds or deletes a single amino acid or nucleotide.

A "ribonucleoprotein complex" as provided herein refers to a complex including a nucleoprotein and a ribonucleic acid. A "nucleoprotein" as provided herein refers to a protein capable of binding a nucleic acid (e.g., RNA, DNA). Where the nucleoprotein binds a ribonucleic acid it is referred to as "ribonucleoprotein." The interaction between the ribonucleoprotein and the ribonucleic acid may be direct, e.g., by covalent bond, or indirect, e.g., by non-covalent bond (e.g. electrostatic interactions (e.g. ionic bond, hydrogen bond, halogen

bond), van der Waals interactions (e.g. dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions and the like). In some embodiments, the ribonucleoprotein includes an RNA-binding motif non-covalently bound to the ribonucleic acid. For example, positively charged aromatic amino acid residues (e.g., lysine residues) in the RNA-binding motif may form electrostatic interactions with the negative nucleic acid phosphate backbones of the RNA, thereby forming a ribonucleoprotein complex. Non-limiting examples of ribonucleoproteins include ribosomes, telomerase, RNaseP, hnRNP, CRISPR associated protein 9 (Cas9) and small nuclear RNPs (snRNPs). The ribonucleoprotein may be an enzyme. In some embodiments, the ribonucleoprotein is an endonuclease.

"Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the nucleic acid or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity over a specified region, e.g., of the entire polypeptide sequences of the present disclosure or individual domains of the polypeptides of the present disclosure), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then to be "substantially identical." This definition also refers to the complement of a test sequence. For example, the one or more homologous donor sequences (e.g., a first and/or a second homologous donor sequence) form part of the donor nucleic acid and may be substantially identical to the DNA targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor

sequence) are 60%, identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 65%, identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 70% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 75% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 80% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 85% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 90% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 95% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 98% identical to the DNA-targeting sequence. In some embodiments, the homologous donor sequences (e.g., a first and/or a second homologous donor sequence) are 99% identical to the DNA-targeting sequence. Optionally, the identity exists over a region that is at least 50 nucleotides in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of, e.g., a full length sequence or from 20 to 600, 50 to 200, or 100 to 150 amino acids or nucleotides in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl.*

Math. 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Ausubel et al., Current Protocols in Molecular Biology* (1995 supplement)).

An example of an algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see Henikoff and Henikoff (1989) Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g., Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)),

which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than 0.2, more preferably less than 0.01, and most preferably less than 0.001.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross-reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

"Biological sample" or "sample" refer to materials obtained from or derived from a subject or patient. A biological sample includes sections of tissues such as biopsy and autopsy samples, and frozen sections taken for histological purposes. Such samples include bodily fluids such as blood and blood fractions or products (e.g., serum, plasma, platelets, red blood cells, and the like), sputum, tissue, cultured cells (e.g., primary cultures, explants, and transformed cells) stool, urine, synovial fluid, joint tissue, synovial tissue, synoviocytes, fibroblast-like synoviocytes, macrophage-like synoviocytes, immune cells, hematopoietic cells, fibroblasts, macrophages, T cells, etc. A biological sample is typically obtained from a eukaryotic organism, such as a mammal such as a primate e.g., chimpanzee or human; cow; dog; cat; a rodent, e.g., guinea pig, rat, mouse; rabbit; or a bird; reptile; or fish.

A "cell" as used herein, refers to a cell carrying out metabolic or other function sufficient to preserve or replicate its genomic DNA. A cell can be identified by methods that include, for example, presence of an intact membrane, staining by a particular dye, ability to produce progeny or, in the case of a gamete, ability to combine with a second gamete to produce a viable offspring. Cells may include prokaryotic and eukaryotic cells. Prokaryotic cells include but are not limited to bacteria. Eukaryotic cells include but are not limited to yeast cells and cells derived from plants and animals, for example mammalian, insect (e.g., spodoptera) and human cells.

The word "expression" or "expressed" as used herein in reference to a gene means the transcriptional and/or translational product of that gene. The level of expression of a DNA molecule in a cell may be determined on the basis of either the amount of corresponding mRNA that is present within the cell or the amount of protein encoded by that DNA produced by the cell (Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*, 18.1-18.88).

Expression of a transfected gene can occur transiently or stably in a cell. During "transient expression" the transfected gene is not transferred to the daughter cell during cell division. Since its expression is restricted to the transfected cell, expression of the gene is lost over time. In contrast, stable expression of a transfected gene can occur when the gene is co-transfected with another gene that confers a selection advantage to the transfected cell. Such a selection advantage may be a resistance towards a certain toxin that is presented to the cell.

The term "exogenous" refers to a molecule or substance (e.g., nucleic acid or protein) that originates from outside a given cell or organism. Conversely, the term "endogenous" refers to a molecule or substance that is native to, or originates within, a given cell or organism.

The terms "transfection", "transduction", "transfecting" or "transducing" can be used interchangeably and are defined as a process of introducing a nucleic acid molecule and/or a protein to a cell. Nucleic acids may be introduced to a cell using non-viral or viral-based methods. The nucleic acid molecule can be a sequence encoding complete proteins or functional portions thereof. Typically, a nucleic acid vector, comprising the elements necessary for protein expression (e.g., a promoter, transcription start site, etc.). Non-viral methods of transfection include any appropriate method that does not use viral DNA or viral particles as a delivery system to introduce the nucleic acid molecule into the cell. Exemplary non-viral transfection methods include calcium phosphate transfection, liposomal transfection, nucleofection, sonoporation, transfection through heat shock, magnetofection and electroporation. For viral-based methods, any useful viral vector can be used in the methods described herein. Examples of viral vectors include, but are not limited to retroviral, adenoviral, lentiviral and adeno-associated viral vectors. In some aspects, the nucleic acid molecules are introduced into a cell using a retroviral vector following standard procedures. The terms "transfection" or "transduction" also refer to introducing proteins into a cell from the external environment. Typically, transduction or transfection of a protein relies on attachment of a peptide or protein capable of crossing the cell membrane to the protein of interest. See, e.g., Ford et al. (2001) *Gene Therapy* 8:1-4 and Prochiantz (2007) *Nat. Methods* 4:119-20.

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are near each other, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The term "gene" means the segment of DNA involved in producing a protein; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons). The leader, the trailer as well as the introns include regulatory elements that are necessary during the transcription and the translation of a gene. Further, a "protein gene product" is a protein expressed from a particular gene.

For specific proteins described herein (e.g., Cas9), the named protein includes any of the protein's naturally occurring forms, or variants or homologs that maintain the protein transcription factor activity (e.g., within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to the native protein). In some embodiments, variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring form. In other embodiments, the protein is the protein as identified by its NCBI sequence reference. In other embodiments, the protein is the protein as identified by its NCBI sequence reference or functional fragment or homolog thereof.

Thus, a "Cas9 nuclease" or "Cas9" protein as referred to herein includes any of the recombinant or naturally-occurring forms of the CRISPR-associated protein 9 (Cas9) or variants or homologs thereof that maintain Cas9 enzyme activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to Cas9). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring Cas9 protein. In some embodiments, the

Cas9 protein is substantially identical to the protein identified by the UniProt reference number Q99ZW2 or a variant or homolog having substantial identity thereto.

"Patient" or "subject in need thereof" refers to a living organism suffering from or prone to a disease or condition that can be treated by administration of a composition or pharmaceutical composition as provided herein. Non-limiting examples include humans, other mammals, bovines, rats, mice, dogs, monkeys, goat, sheep, cows, deer, and other non-mammalian animals. In some embodiments, a patient is human.

Effective doses of the compositions provided herein (e.g., the ribonucleoprotein complex, the DNA repair protein (e.g., conjugate) or nucleic acids encoding the same) may be administered to a subject in need thereof for treating and preventing a disease (e.g., cancer).

The terms "disease" or "condition" refer to a state of being or health status of a patient or subject capable of being treated with a compound, pharmaceutical composition, or method provided herein. In some embodiments, the disease is cancer (e.g. lung cancer, ovarian cancer, osteosarcoma, bladder cancer, cervical cancer, liver cancer, kidney cancer, skin cancer (e.g., Merkel cell carcinoma), testicular cancer, leukemia, lymphoma (Mantel cell lymphoma), head and neck cancer, colorectal cancer, prostate cancer, pancreatic cancer, melanoma, breast cancer, neuroblastoma).

As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably herein. These terms refer to an approach for obtaining beneficial or desired results including but not limited to therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant eradication or amelioration of the underlying disorder being treated. Also, a therapeutic benefit is achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disorder such that an improvement is observed in the patient, notwithstanding that the patient may still be afflicted with the underlying disorder. For prophylactic benefit, the compositions may be administered to a patient at risk of developing a particular disease, or to a patient reporting one or more of the physiological symptoms of a disease, even though a diagnosis of this disease may not have been made. Treatment includes preventing the disease, that is, causing the clinical symptoms of the disease not to develop by administration of a protective composition prior to the induction of the disease; suppressing the disease, that is, causing the clinical symptoms of the disease not to develop by administration of a protective composition after the inductive event but prior to the clinical appearance or reappearance of the disease; inhibiting the disease, that is, arresting the development of clinical symptoms by administration of a protective composition after their

initial appearance; preventing re-occurring of the disease and/or relieving the disease, that is, causing the regression of clinical symptoms by administration of a protective composition after their initial appearance.

An "effective amount" is an amount sufficient to accomplish a stated purpose (e.g. achieve the effect for which it is administered, treat a disease, reduce enzyme activity, reduce one or more symptoms of a disease or condition). An example of an "effective amount" is an amount sufficient to contribute to the treatment, prevention, or reduction of a symptom or symptoms of a disease, which could also be referred to as a "therapeutically effective amount." A "reduction" of a symptom or symptoms (and grammatical equivalents of this phrase) means decreasing of the severity or frequency of the symptom(s), or elimination of the symptom(s). A "prophylactically effective amount" of a drug is an amount of a drug that, when administered to a subject, will have the intended prophylactic effect, e.g., preventing or delaying the onset (or reoccurrence) of an injury, disease, pathology or condition, or reducing the likelihood of the onset (or reoccurrence) of an injury, disease, pathology, or condition, or their symptoms. The full prophylactic effect does not necessarily occur by administration of one dose, and may occur only after administration of a series of doses. Thus, a prophylactically effective amount may be administered in one or more administrations. An "activity decreasing amount," as used herein, refers to an amount of antagonist required to decrease the activity of an enzyme or protein relative to the absence of the antagonist. A "function disrupting amount," as used herein, refers to the amount of antagonist required to disrupt the function of an enzyme or protein relative to the absence of the antagonist. Guidance can be found in the literature for appropriate dosages for given classes of pharmaceutical products. For example, for the given parameter, an effective amount will show an increase or decrease of at least 5%, 10%, 15%, 20%, 25%, 40%, 50%, 60%, 75%, 80%, 90%, or at least 100%. Efficacy can also be expressed as "-fold" increase or decrease. For example, a therapeutically effective amount can have at least a 1.2-fold, 1.5-fold, 2-fold, 5-fold, or more effect over a control. The exact amounts will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using various techniques (*see, e.g.,* Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); Pickar, *Dosage Calculations* (1999); and *Remington: The Science and Practice of Pharmacy*, 20th Edition, 2003, Gennaro, Ed., Lippincott, Williams & Wilkins).

The compounds of the present disclosure can be administered alone or can be co-administered to the patient. Co-administration is meant to include simultaneous or sequential

administration of the compounds individually or in combination (more than one compound). Thus, the preparations can also be combined, when desired, with other active substances (e.g. to reduce metabolic degradation).

The combined administration contemplates co-administration, using separate formulations or a single pharmaceutical formulation, and consecutive administration in either order, wherein preferably there is a time period while both (or all) active agents simultaneously exert their biological activities.

“Pharmaceutically acceptable excipient” and “pharmaceutically acceptable carrier” refer to a substance that aids the administration of an active agent to and absorption by a subject and can be included in the compositions of the present disclosure without causing a significant adverse toxicological effect on the patient.

Target Nucleic Acids

The methods, compositions, and kits of the present disclosure may be used to, in some embodiments, to modify a target nucleic acid sequence. A "target nucleic acid sequence" may be any nucleic acid sequence modified as provided herein (e.g., to which a programmable nuclease and/or DNA repair protein is localized). A target nucleic acid sequence may include a site that is hydrolyzed (cleaved) by a programmable nuclease (e.g., a RNA-guided nuclease, such as Cas9, a ZFN, or a TALEN). Thus, in some embodiments, a target nucleic acid sequence includes an nuclease cleavage site.

In some embodiments, a target nucleic acid sequence is an exogenous nucleic acid sequence. In some embodiments, a target nucleic acid sequence is an endogenous nucleic acid sequence. In some embodiments, a target nucleic acid sequence forms part of a cellular, e.g., genomic, gene. In some embodiments, a target nucleic acid sequence is part of a transcriptional regulatory sequence. In some embodiments, a target nucleic acid sequence is part of a promoter, enhancer or silencer.

In some embodiments, a target sequence is a DNA. In other embodiments, a target sequence is a RNA.

In some embodiments, a target nucleic acid sequence is at, near, or within a promoter sequence. In some embodiments, a target nucleic acid sequence is at, near, or within a gene. In some embodiments, a target nucleic acid sequence is known to be associated with a disease or condition characterized by a (one or more) nucleotide mutation (e.g., substitution), insertion or deletion. In some embodiments, a target nucleic acid sequence is within a tumor suppressor gene

or an oncogene, such as within a transcriptional regulatory sequence/element or coding region of the tumor suppressor gene or oncogene.

In some embodiments, a target nucleic acid sequence is immediately 3' to a protospacer adjacent motif (PAM) sequence. In some embodiments, a PAM sequence of a target nucleic acid sequence is 5' -CCN-3', wherein N is any DNA nucleotide. In some embodiments, a PAM sequence of a target nucleic acid sequence matches the Cas9 endonuclease binding site or Cas9 nickase binding site or homologs or orthologs to be used.

For Cas9 endonuclease or Cas9 nickase to successfully bind to DNA, the target nucleic acid sequence in the genomic DNA should be complementary to a guide RNA sequence and immediately followed by a correct PAM sequence. A PAM sequence is present in the target nucleic acid sequence but not in the guide RNA sequence. Any DNA sequence with the correct target nucleic acid sequence followed by a PAM sequence should be bound by Cas9 endonuclease or Cas9 nickase. A PAM sequence may be any of the PAM sequences disclosed in international application PCT/US2016/021491 and published as WO2016148994 A8, which is hereby incorporated by reference. Other PAM sequences are known and may be used herein.

DNA Repair Complexes

The DNA repair complexes provided herein are, *inter alia*, useful for editing genome sequences by introducing precise changes in a target site in the presence of a donor sequence. The RNA-guided DNA endonuclease provided herein including embodiments thereof (e.g., Cas9 nuclease or Cas9 nickase) is capable of introducing a strand break (double- or single-strand break) at a target site in the genome of a cell (e.g., gene or transcriptional regulatory sequence) and the break is then predominantly repaired through the mechanism of HDR. By increasing the HDR efficiency significantly and in certain instances decreasing NHEJ at a target site, the compositions and methods provided herein meet the long-felt need of site directed, highly accurate genome editing. The compositions provided herein including embodiments thereof are therefore widely useful as therapeutics and research tools. Effective doses of the RNA-guided DNA endonuclease, the nucleic acid (e.g., guide RNA), the DNA repair protein (e.g., conjugate) and the donor nucleic acid provided herein as well as nucleic acids encoding the same may be administered to a subject in need thereof for treating and preventing a disease (e.g., cancer).

The DNA repair complexes provided herein including embodiments thereof are based on a three-component hybrid system (also known as *Casilio* system). The *Casilio* system includes CRISPR/Cas9, guide RNA including PBS and PUF domain coupled with the DNA repair protein

and Pumilio proteins. For purpose of the present disclosure, the three-component hybrid system that includes CRISPR/Cas9 and Pumilio proteins may also be referred to interchangeably as the *Casilio* system. The *Casilio* system is used for the targeted delivery of DNA repair protein domains (e.g., DNA ligases, nucleases, helicases) to a specific site in the genome. In essence, the DNA repair protein domain is linked to (e.g., fused) to the N-terminus or the C-terminus of Pumilio proteins or functional fragments thereof (PUF domains) that bind PBS in the *Casilio* system, thus bringing such domains to the vicinity of any target locus of interest that is specifically recognized by the *Casilio* system. Any aspects or embodiments of the three-component CRISPR/Cas complex system disclosed in international application PCT/US2016/021491 and published as WO2016148994 A8, which is hereby incorporated by reference and for all purposes, may be used as provided herein.

The compositions and methods provided herein including embodiments thereof are advantageous over previous attempts to edit a target gene sequence in a cell using programmable nuclease. The present disclosure, in some embodiments, permits the precise editing at specific locations in the genome, for example, by increasing HDR at a target site.

In one aspect, a DNA repair complex is provided. The complex includes: (a) a ribonucleoprotein complex including: (i) an RNA-guided DNA endonuclease; and (ii) a nucleic acid including: (1) a DNA-targeting sequence that is complementary to a target nucleic acid sequence; (2) a binding sequence for the RNA-guided DNA endonuclease; and (3) one or more PUF binding site (PBS) sequences, wherein the RNA-guided DNA endonuclease is bound to the nucleic acid via the binding sequence; and (b) a DNA repair protein (e.g., conjugate) including: (i) a PUF domain, the PUF domain having a C-terminus and an N-terminus; and (ii) a DNA repair domain, the DNA repair domain linked to the PUF domain to form a DNA repair protein, wherein the DNA repair protein binds to the ribonucleoprotein complex via the PUF domain binding to the one or more PBS sequences to form a DNA repair complex, and wherein when the RNA-guided DNA endonuclease introduces a strand break at a target nucleic acid sequence present in a genome, the DNA repair protein repairs the strand break favoring homology-directed repair (HDR).

Ribonucleoprotein complexes

RNA-Guided DNA Endonucleases

The ribonucleoprotein complexes provided herein including embodiments thereof include an RNA-guided DNA endonuclease bound through a binding sequence to a nucleic acid (e.g.,

guide RNA). The nucleic acid further includes a DNA targeting sequence, which is complementary to a target nucleic acid sequence in the genome, and one or more PUF binding site (PBS) sequences.

For the compositions and methods provided herein, any RNA-guided DNA endonuclease may be used. An “RNA-guided DNA endonuclease” as provided herein refers to an endonuclease that can be recruited to a target sequence in the genome by a guide RNA and which is capable of introducing a strand break at a target sequence. The DNA nuclease binds the guide RNA and the guide RNA is capable of hybridizing to a target sequence. The RNA-guided DNA endonuclease provided herein may introduce a single strand break or a double strand break at a target nucleic acid sequence present in a genome. Non-limiting examples of RNA-guided DNA endonucleases include Cas9 nuclease, Cas9 nickase. Thus, in some embodiments, the RNA-guided DNA endonuclease is a Cas9 nuclease. Where the RNA-guided DNA endonuclease is a Cas9 nuclease, the endonuclease may introduce a double strand-break at a target nucleic acid sequence (i.e., a break at sense strand and a break at the antisense strand). Where the RNA-guided DNA endonuclease is a Cas9 nickase, the endonuclease may introduce a single strand-break at a target nucleic acid sequence (i.e., a break at sense strand or a break at the antisense strand).

In some embodiments, the RNA-guided DNA endonuclease is a Cas9 nickase. In some embodiments, the RNA-guided DNA endonuclease includes an alanine at a position corresponding to amino acid position 10 of SEQ ID NO: 25. In some embodiments, the RNA-guided DNA endonuclease is a Cas9 D10A nickase. In some embodiments, the RNA-guided DNA endonuclease includes an alanine corresponding to amino acid position 840 of SEQ ID NO: 25. In some embodiments, the RNA-guided DNA endonuclease is a Cas9 H840A nickase.

In some embodiments, the RNA-guided DNA endonuclease is SpCas9 from *S. Pyrogenes*. In some embodiments, the RNA-guided DNA endonuclease is Clustered Regularly Interspaced Short Palindromic Repeats from *Prevotella* and *Francisella* 1 (Cpf1). In some embodiments, the Cas9 endonuclease includes the sequence of SEQ ID NO: 25. In some embodiments, the RNA-guided DNA endonuclease is the sequence of SEQ ID NO: 25. In some embodiments, the RNA-guided DNA endonuclease includes the sequence of SEQ ID NO: 26. In some embodiments, the RNA-guided DNA endonuclease is the sequence of SEQ ID NO: 26. In some embodiments, the RNA-guided DNA endonuclease includes the sequence of SEQ ID NO: 89. In some embodiments, the RNA-guided DNA endonuclease is the sequence of SEQ ID NO: 89.

Any of the endonucleases described in Leinstiver et al. (Nature, vol. 529, pages 490–495 (28 January 2016) and Hu et al. (Nature, vol 556, pages 57–63 (05 April 2018)), which are hereby incorporated by reference in their entirety and for all purposes, may be used for the compositions and methods proved herein.

Guide RNAs and Other Nucleic Acids

A nucleic acid provided herein, in some embodiments, includes (1) a DNA-targeting sequence that is complementary to a target nucleic acid sequence, (2) a binding sequence for the RNA-guided DNA endonuclease (e.g., Cas9 nuclease, Cas9 nickase, Cas9 H840A nickase, Cas9 D10A nickase), and (3) one or more PUF binding site (PBS) sequences. In some embodiments, the complex includes Cas9 nuclease bound to the nucleic acid thereby forming a ribonucleoprotein complex. In some embodiments, the complex includes Cas9 nickase bound to the nucleic acid thereby forming a ribonucleoprotein complex. In some embodiments, the nucleic acid is a ribonucleic acid. In some embodiments, the nucleic acid is a guide RNA. A "guide RNA" or "gRNA" as provided herein refers to a ribonucleotide sequence capable of binding a nucleoprotein, thereby forming ribonucleoprotein complex. In some embodiments, the nucleic acid of the present disclosure can be a single RNA molecule (single RNA nucleic acid), which may include a "single-guide RNA" (abbreviated to "sgRNA" or "gRNA" In another embodiment, the nucleic acid of the present disclosure includes two RNA molecules (e.g., joined together via hybridization at the binding sequence (e.g., Cas9 nuclease-binding sequence). Thus the subject nucleic acid is inclusive, referring both to two-molecule nucleic acids and to single-molecule nucleic acids (e.g., sgRNAs).

In some embodiments, the nucleic acid (e.g., gRNA) is a single-stranded ribonucleic acid. In some embodiments, the nucleic acid (e.g., gRNA) is 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more nucleic acid residues in length. In some embodiments, the nucleic acid (e.g., gRNA) is from 10 to 30 nucleic acid residues in length. In some embodiments, the nucleic acid (e.g., gRNA) is 20 nucleic acid residues in length. In some embodiments, the length of the nucleic acid (e.g., gRNA) can be at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more nucleic acid residues or sugar residues in length. In some embodiments, the nucleic acid (e.g., gRNA) is from 5 to 50, 10 to 50, 15 to 50, 20 to 50, 25 to 50, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 5 to 75, 10 to 75, 15 to 75, 20 to 75, 25 to 75, 30 to 75, 35 to 75, 40 to 75, 45 to 75, 50

to 75, 55 to 75, 60 to 75, 65 to 75, 70 to 75, 5 to 100, 10 to 100, 15 to 100, 20 to 100, 25 to 100, 30 to 100, 35 to 100, 40 to 100, 45 to 100, 50 to 100, 55 to 100, 60 to 100, 65 to 100, 70 to 100, 75 to 100, 80 to 100, 85 to 100, 90 to 100, 95 to 100, or more residues in length. In some embodiments, the nucleic acid (e.g., gRNA) is from 10 to 15, 10 to 20, 10 to 30, 10 to 40, or 10 to 50 residues in length.

In some embodiments, transcription of the nucleic acid is under the control of a constitutive promoter, such as a CMV promoter or a Ubc promoter, or an inducible promoter, such as a tetracycline-responsive promoter or a steroid-responsive promoter. In some embodiments, the nucleic acid is a vector. In some embodiments, transcription of the nucleic acid is under the control of an RNA promoter. In some embodiments, the RNA promoter is a U6 promoter. In some embodiments, the RNA promoter is an H1 promoter.

In some embodiments, the vector encoding the nucleic acid (for use in the methods of the present disclosure) is active in a cell from a mammal (a human; a non-human primate; a non-human mammal; a rodent); an insect, a worm, a yeast, or a bacterium.

In some embodiments, the vector is a plasmid, a viral vector (such as adenoviral, retroviral, or lentiviral vector, or AAV vector), or a transposon (such as piggy Bac transposon). The vector can be transiently transfected into a host cell, or be integrated into a host genome by infection or transposition.

The present disclosure also provides an isolated nucleic acid comprising a nucleotide sequence encoding a gRNA. In some embodiments, a nucleic acid also comprises a nucleotide sequence encoding a RNA-guided DNA endonuclease (Cas9 protein) and/or a DNA repair protein (e.g., conjugate).

DNA-Targeting Sequences

The nucleic acid includes a nucleotide sequence complementary to a target site (e.g., target nucleic acid sequence), which is referred to herein as "DNA-targeting sequence." The DNA-targeting sequence may mediate binding of the ribonucleoprotein complex to a complementary target nucleic acid sequence thereby providing the sequence specificity of the ribonucleoprotein complex. Thus, In some embodiments, the nucleic acid (e.g., gRNA) or parts thereof are complementary to a target nucleic acid sequence. In some embodiments, the nucleic acid (e.g., gRNA) binds a target nucleic acid sequence. In some embodiments, the complement of the nucleic acid has a sequence identity of 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% to a target nucleic acid sequence. In some embodiments, the

complement of the DNA-targeting sequence has a sequence identity of 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% to a target nucleic acid sequence.

It should be noted that the DNA-targeting sequence may or may not be 100% complementary to a target nucleic acid sequence. In some embodiments, the DNA-targeting sequence is complementary to a target nucleic acid sequence over 8-25 nucleotides (nts), 12-22 nucleotides, 14-20 nts, 16-20 nts, 18-20 nts, or 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nts. In some embodiments, the complementary region includes a continuous stretch of 12-22 nts, preferably at the 3' end of the DNA-targeting sequence. In some embodiments, the 5' end of the DNA-targeting sequence has up to 8 nucleotide mismatches with a target nucleic acid sequence. In some embodiments, the DNA-binding sequence is 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100% complementary to a target nucleic acid sequence.

In a related embodiment, there is no more than 15-nucleotide match at the 3' end of the DNA-targeting sequence compared to the complementary target nucleic acid sequence, and the RNA-guided DNA endonuclease in the complex is a wildtype Cas9 protein. In some embodiments, the RNA-guided DNA endonuclease is a Cas9 nickase. In some embodiments, the RNA-guided DNA endonuclease is a Clustered Regularly Interspaced Short Palindromic Repeats from *Prevotella* and *Francisella* 1 (Cpf1).

The DNA-targeting sequence is functionally similar or equivalent to the crRNA or guide RNA or gRNA of the CRISPR/Cas complex / system. However, in the context of the instant disclosure, the DNA-targeting sequence may not originate from any particular crRNA or gRNA, but can be arbitrarily designed based on the sequence of a target nucleic acid sequence.

The DNA-targeting sequence includes a nucleotide sequence that is complementary to a specific sequence within a target DNA (or the complementary strand of a target DNA). In other words, the DNA-targeting sequence interacts with a target nucleic acid sequence of a target DNA in a sequence-specific manner via hybridization (*i.e.*, base pairing). As such, the nucleotide sequence of the DNA-targeting sequence may vary, and it determines the location within a target DNA that the subject nucleic acid and a target DNA will interact. The DNA-targeting sequence can be modified or designed (*e.g.*, by genetic engineering) to hybridize to any desired sequence within a target DNA. In some embodiments, a target nucleic acid sequence is immediately 3' to a PAM (protospacer adjacent motif) sequence of the complementary strand, which can be 5' - CCN-3' , wherein N is any DNA nucleotide. That is, in this embodiment, the complementary strand of a target nucleic acid sequence is immediately 5' to a PAM sequence that is 5'-NGG-3', wherein N is any DNA nucleotide.

The DNA-targeting sequence can have a length of from 12 nucleotides to 100 nucleotides. For example, the DNA-targeting sequence can have a length of from 12 nucleotides (nt) to 80 nt, from 12 nt to 50 nt, from 12 nt to 40 nt, from 12 nt to 30 nt, from 12 nt to 25 nt, from 12 nt to 20 nt, or from 12 nt to 19 nt. For example, the DNA-targeting sequence can have a length of from 19 nt to 20 nt, from 19 nt to 25 nt, from 19 nt to 30 nt, from 19 nt to 35 nt, from 19 nt to 40 nt, from 19 nt to 45 nt, from 19 nt to 50 nt, from 19 nt to 60 nt, from 19 nt to 70 nt, from 19 nt to 80 nt, from 19 nt to 90 nt, from 19 nt to 100 nt, from 20 nt to 25 nt, from 20 nt to 30 nt, from 20 nt to 35 nt, from 20 nt to 40 nt, from 20 nt to 45 nt, from 20 nt to 50 nt, from 20 nt to 60 nt, from 20 nt to 70 nt, from 20 nt to 80 nt, from 20 nt to 90 nt, or from 20 nt to 100 nt.

The nucleotide sequence of the DNA-targeting sequence that is complementary to a target nucleic acid sequence of a target DNA can have a length of at least 12 nt. For example, the DNA-targeting sequence that is complementary to a target nucleic acid sequence of a target DNA can have a length at least 12 nt, at least 15 nt, at least 18 nt, at least 19 nt, at least 20 nt, at least 25 nt, at least 30 nt, at least 35 nt or at least 40 nt. For example, the DNA-targeting sequence that is complementary to a target nucleic acid sequence of a target DNA can have a length of from 12 nucleotides (nt) to 80 nt, from 12 nt to 50 nt, from 12 nt to 45 nt, from 12 nt to 40 nt, from 12 nt to 35 nt, from 12 nt to 30 nt, from 12 nt to 25 nt, from 12 nt to 20 nt, from 12 nt to 19 nt, from 19 nt to 20 nt, from 19 nt to 25 nt, from 19 nt to 30 nt, from 19 nt to 35 nt, from 19 nt to 40 nt, from 19 nt to 45 nt, from 19 nt to 50 nt, from 19 nt to 60 nt, from 20 nt to 25 nt, from 20 nt to 30 nt, from 20 nt to 35 nt, from 20 nt to 40 nt, from 20 nt to 45 nt, from 20 nt to 50 nt, or from 20 nt to 60 nt. The nucleotide sequence of the DNA-targeting sequence that is complementary to a target nucleic acid sequence of a target DNA can have a length of at least 12 nt.

In some cases, the DNA-targeting sequence that is complementary to a target nucleic acid sequence of a target DNA is 20 nucleotides in length. In some cases, the DNA-targeting sequence that is complementary to a target nucleic acid sequence of a target DNA is 19 nucleotides in length.

The percent complementarity between the DNA-targeting sequence and a target nucleic acid sequence can be at 50% (*e.g.*, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100%). In some cases, the percent complementarity between the DNA-targeting sequence and a target nucleic acid sequence is 100% over the seven or eight contiguous 5'-most nucleotides of a target nucleic acid sequence. In some embodiments, the percent

complementarity between a DNA-targeting sequence and a target nucleic acid sequence is at least 60% over 20 contiguous nucleotides. In some embodiments, the percent complementarity between the DNA-targeting sequence and a target nucleic acid sequence is 100% over the 7, 8, 9, 10, 11, 12, 13, or 14 contiguous 5'-most nucleotides of a target nucleic acid sequence (*i.e.*, the 7, 8, 9, 10, 11, 12, 13, or 14 contiguous 3'-most nucleotides of the DNA-targeting sequence), and as low as 0% over the remainder. In such a case, the DNA-targeting sequence can be considered to be 7, 8, 9, 10, 11, 12, 13, or 14 nucleotides in length, respectively.

In some embodiments, the nucleic acid (*e.g.*, gRNA) is 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% to a target nucleic acid sequence. In some embodiments, the nucleic acid (*e.g.*, gRNA) is 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% complementary to the sequence of a cellular gene. In some embodiments, the nucleic acid (*e.g.*, gRNA) binds a cellular gene sequence.

RNA-Guided Nuclease-Binding Sequences

In some embodiments, the complex includes a Cas9 endonuclease or a Cas9 nickase bound to the nucleic acid through binding a binding sequence of the nucleic acid and thereby forming a ribonucleoprotein complex. In some embodiments, the binding sequence forms a hairpin structure. In some embodiments, the binding sequence is 30-100 nt, 35-50 nt, 37-47 nt, or 42 nt in length. In some embodiments, the binding sequence is the sequence of SEQ ID NO: 89. In some embodiments, the binding sequence is the sequence of SEQ ID NO: 90. In some embodiments, the binding sequence includes the sequence of SEQ ID NO: 89. In some embodiments, the binding sequence includes the sequence of SEQ ID NO:90.

The binding sequence (protein-binding segment or protein-binding sequence) of the subject nucleic acid binds to a RNA-guided DNA endonuclease (*e.g.*, Cas9 endonuclease or a Cas9 nickase). For simplicity, the binding sequence (protein-binding segment or protein-binding sequence), which may bind to a RNA-guided DNA endonuclease (*e.g.*, Cas9 endonuclease or a Cas9 nickase), may simply be referred to as "Cas9-binding sequence" or "binding sequence" herein.

The binding sequence (Cas9-binding sequence) interacts with or is bound by a RNA-guided DNA endonuclease (*e.g.*, Cas9 endonuclease or a Cas9 nickase), and together they bind to a target nucleic acid sequence recognized by the DNA-targeting sequence. The binding sequence (Cas9-binding sequence) includes two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (a dsRNA duplex). These two

complementary stretches of nucleotides may be covalently linked by intervening nucleotides known as linkers or linker nucleotides (*e.g.*, in the case of a single-molecule nucleic acid), and hybridize to form the double stranded RNA duplex (dsRNA duplex, or “Cas9-binding hairpin”) of the binding sequence (Cas9-binding sequence), thus resulting in a stem-loop structure.

Alternatively, in some embodiment, the two complementary stretches of nucleotides may not be covalently linked, but instead are held together by hybridization between complementary sequences (*e.g.*, in the case of a two-molecule nucleic acid of the present disclosure).

The binding sequence (Cas9-binding sequence) can have a length of from 10 nucleotides to 100 nucleotides, *e.g.*, from 10 nucleotides (nt) to 20 nt, from 20 nt to 30 nt, from 30 nt to 40 nt, from 40 nt to 50 nt, from 50 nt to 60 nt, from 60 nt to 70 nt, from 70 nt to 80 nt, from 80 nt to 90 nt, or from 90 nt to 100 nt. For example, the Cas9-binding sequence can have a length of from 15 nucleotides (nt) to 80 nt, from 15 nt to 50 nt, from 15 nt to 40 nt, from 15 nt to 30 nt, from 37 nt to 47 nt (*e.g.*, 42 nt), or from 15 nt to 25 nt.

The dsRNA duplex of the binding sequence (Cas9-binding sequence) can have a length from 6 base pairs (bp) to 50 bp. For example, the dsRNA duplex of the binding sequence (Cas9-binding sequence) can have a length from 6 bp to 40 bp, from 6 bp to 30 bp, from 6 bp to 25 bp, from 6 bp to 20 bp, from 6 bp to 15 bp, from 8 bp to 40 bp, from 8 bp to 30 bp, from 8 bp to 25 bp, from 8 bp to 20 bp or from 8 bp to 15 bp. For example, the dsRNA duplex of the binding sequence (Cas9-binding sequence) can have a length from 8 bp to 10 bp, from 10 bp to 15 bp, from 15 bp to 18 bp, from 18 bp to 20 bp, from 20 bp to 25 bp, from 25 bp to 30 bp, from 30 bp to 35 bp, from 35 bp to 40 bp, or from 40 bp to 50 bp. In some embodiments, the dsRNA duplex of the binding sequence (Cas9-binding sequence) has a length of 36 base pairs. The percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the binding sequence (Cas9-binding sequence) can be at least 60%. For example, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the binding sequence (Cas9-binding sequence) can be at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99%. In some embodiments, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the binding sequence (Cas9-binding sequence) is 100%.

In some embodiments, the nucleic acid further includes a linker sequence linking the DNA-targeting sequence to the binding sequence (Cas9-binding sequence). The linker can have a length of from 3 nucleotides to 100 nucleotides. For example, the linker can have a length of 3 nucleotides (nt) to 90 nt, from 3 nucleotides (nt) to 80 nt, from 3 nucleotides (nt) to 70 nt, from 3

nucleotides (nt) to 60 nt, from 3 nucleotides (nt) to 50 nt, from 3 nucleotides (nt) to 40 nt, from 3 nucleotides (nt) to 30 nt, from 3 nucleotides (nt) to 20 nt or from 3 nucleotides (nt) to 10 nt. For example, the linker can have a length of from 3 nt to 5 nt, from 5 nt to 10 nt, from 10 nt to 15 nt, from 15 nt to 20 nt, from 20 nt to 25 nt, from 25 nt to 30 nt, from 30 nt to 35 nt, from 35 nt to 40 nt, from 40 nt to 50 nt, from 50 nt to 60 nt, from 60 nt to 70 nt, from 70 nt to 80 nt, from 80 nt to 90 nt, or from 90 nt to 100 nt. In some embodiments, the linker is 4 nt.

Non-limiting examples of nucleotide sequences that can be included in a suitable binding sequence (Cas9-binding sequence, *i.e.*, Cas9 handle) are set forth in SEQ ID NOs: 563-682 of WO 2013/176772 (see, for examples, FIGs. 8 and 9 of WO 2013/176772), which is hereby incorporated by reference in its entirety and for all purposes.

In some embodiments, the binding sequence (Cas9-binding sequence) includes a nucleotide sequence that differs by 1, 2, 3, 4, or 5 nucleotides from any one of the above-listed sequences. In some embodiments, the binding sequence (Cas9-binding sequence) includes a nucleotide sequence that has 98%, 97%, 96% or 95% sequence identity to any one of the above-listed sequences.

PUF Domain Binding Sequences

The term “PBS” sequences or “PUF binding site” sequences as provided herein refers to a site that is bound by a Pumilio/fem-3 mRNA binding factor (PUF). A PUF binding site (PBS) may form part of a guide RNA and provide for the binding of a PUF protein or PUF domain as provided herein (e.g., PUFa, PUFb, PUFc or functional fragments thereof) to the guide RNA. The PUF binding site includes a nucleic acid sequence (*i.e.*, a PBS sequence or PUF binding site sequence) which is characteristic of the PBS and may be bound directly by the PUF protein. The nucleic acid (e.g., gRNA) provided herein further includes one or more PUF binding site (PBS) sequences. In some embodiments, the one or more PBS sequences contain 8 nucleotides in length. In some embodiments, the one or more PBS sequences contain at least 9 nucleotides in length. In some embodiments, the one or more PBS sequences contain at 10 nucleotides in length. In some embodiments, the one or more PBS sequences contain at 11 nucleotides in length. In some embodiments, the one or more PBS sequences contain at 12 nucleotides in length. In some embodiments, the one or more PBS sequences contain at 13 nucleotides in length. In some embodiments, the one or more PBS sequences contain at 14 nucleotides in length. In some embodiments, the one or more PBS sequences contain at 15 nucleotides in length. In some embodiments, the one or more PBS sequences contain at 16 nucleotides in length.

Any of the PBS sequences disclosed in Katarzyna et al. (PNAS May 10, 2016. 113 (19) E2579-E2588) and Zhao et al. (Nucleic Acids Research, Volume 46, Issue 9, 18 May 2018, Pages 4771–4782) may be used as provided herein including embodiments thereof.

In some embodiments, the one or more PBS sequences are identical. In some embodiments, the nucleic acid includes 1 to 50 PBS sequences. Any one of the PBS sequences disclosed in international application PCT/US2016/021491 and published as WO2016148994 A8, which is hereby incorporated by reference in its entirety and for all purposes, are contemplated for the compositions and methods provided herein.

The nucleic acid of the present disclosure may have more than one copy of the PBS sequences. In some embodiments, the nucleic acid comprises 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 46, 47, 48, 49, or 50 copies of PBS sequences, such as 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 copies of PBS sequences. In some embodiments, the range of the PBS sequence copy number is L to H, wherein L is any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, or 40, and wherein H is any one of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, or 100, so long as H is greater than L. Each PBS sequence may be the same or different.

In some embodiments, a nucleic acid includes 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 46, 47, 48, 49, or 50 copies, or 1-50, 2-45, 3-40, 5-35, 5-10, 10-20 copies of identical or different PBS sequences.

In some embodiments, the nucleic acid includes 5-15 copies of PBS sequences, or 5-14 copies, 5-13 copies, 5-12 copies, 5-11 copies, 5-10 copies, or 5-9 copies of PBS sequences.

In some embodiments, the amount of the gRNA-PBS sequences and/or the amount of the DNA repair protein (e.g., conjugate) transfected or expressed is adjusted to maximize PBS/PUF domain binding. For example, this can be achieved by increasing the expression of the PUF domain by a stronger promoter or using an inducible promoter, such as a Dox-inducible promoter.

In some embodiments, the spacing between PBS sequences and/or spacer sequences are optimized to improve system efficiency. For example, spacing optimization can be subject to particular DNA repair proteins (e.g., conjugates), and can be different between proteins that work as individual proteins and those DNA repair protein that may need to be positioned close enough to function (e.g., protein complexes).

In some embodiments, one or more spacer region(s) separate two adjacent PBS sequences. The spacer regions may have a length of from 3 nucleotides to 100 nucleotides. For

example, the spacer can have a length of from 3 nucleotides (nt) to 90 nt, from 3 nucleotides (nt) to 80 nt, from 3 nucleotides (nt) to 70 nt, from 3 nucleotides (nt) to 60 nt, from 3 nucleotides (nt) to 50 nt, from 3 nucleotides (nt) to 40 nt, from 3 nucleotides (nt) to 30 nt, from 3 nucleotides (nt) to 20 nt or from 3 nucleotides (nt) to 10 nt. For example, the spacer can have a length of from 3 nt to 5 nt, from 5 nt to 10 nt, from 10 nt to 15 nt, from 15 nt to 20 nt, from 20 nt to 25 nt, from 25 nt to 30 nt, from 30 nt to 35 nt, from 35 nt to 40 nt, from 40 nt to 50 nt, from 50 nt to 60 nt, from 60 nt to 70 nt, from 70 nt to 80 nt, from 80 nt to 90 nt, or from 90 nt to 100 nt. In some embodiments, the spacer is 4 nt.

In some embodiments, the one or more PBS sequences contain 8 nucleotides in length. In some embodiments, the one or more PBS sequences are identical. In some embodiments, the nucleic acid includes 1 to 50 PBS sequences. In some embodiments, the one or more PBS sequences include the nucleotide sequence of SEQ ID NO: 83.

DNA Repair Proteins

The DNA repair proteins (e.g., conjugates) provided herein are, *inter alia*, useful for repairing the strand break introduced by the RNA-guided DNA endonuclease provided herein. The DNA repair proteins provided herein include two domains, a PUF domain capable of binding the one or more PBS sequences, and a DNA repair domain, which is linked to the PUF domain and mediates repair of the strand break in a target nucleic acid sequence. Through binding of the PUF domain to the one or more PBS sequences included in the nucleic acid, the DNA repair proteins provided herein are recruited to a target nucleic acid sequence in the genome of a cell. Upon recruitment to a target nucleic acid sequence and introduction of a strand break at a target nucleic acid sequence by the RNA-guided DNA endonuclease, the DNA repair domain subsequently repairs the strand break.

PUF Domains

PUF proteins (named after *Drosophila* Pumilio and *C. elegans* fem-3 binding factor) are involved in mediating mRNA stability and translation. These proteins contain a unique RNA-binding domain known as the PUF domain. The RNA-binding PUF domain, such as that of the human Pumilio 1 protein (referred here also as PUM), contains 8 repeats (each repeat called a PUF motif or a PUF repeat) that bind consecutive bases in an anti-parallel fashion, with each repeat recognizing a single base - *i.e.*, PUF repeats R1 to R8 recognize nucleotides N8 to N1, respectively. For example, PUM is composed of eight tandem repeats, each repeat consisting of 34 amino acids that folds into tightly packed domains composed of alpha helices. In some embodiments, the PUF domain binds 8, 9 or 16 nucleotides of the PUF binding site (PBS)

sequence. In some embodiments, the PUF domain is any of the domains disclosed in Katarzyna et al. (PNAS May 10, 2016. 113 (19) E2579-E2588) or Zhao et al. (Nucleic Acids Research, Volume 46, Issue 9, 18 May 2018, Pages 4771–4782), which are herewith incorporated by reference in their entirety and for all purposes.

The DNA repair proteins provided herein including embodiments thereof may be proteins (e.g., conjugates) that include a PUF domain linked to a DNA repair domain. The DNA repair domain may be linked to the N-terminus or the C-terminus of the PUF domain.

As used herein, the term “PUF domain” refers to a wildtype or naturally existing PUF domain, as well as a PUF homologue domain that is based on / derived from a natural or existing PUF domain, such as the prototype human Pumilio 1 PUF domain. The PUF domain of the present disclosure specifically binds to an RNA sequence (e.g., an 8-mer RNA sequence), wherein the overall binding specificity between the PUF domain and the RNA sequence is defined by sequence specific binding between each PUF motif / PUF repeat within the PUF domain and the corresponding single RNA nucleotide.

Also included in the scope of the present disclosure are functional variants of the subject PUF domains or fusions thereof. The term “functional variant” as used herein refers to a PUF domain having substantial or significant sequence identity or similarity to a parent PUF domain, which functional variant retains the biological activity of the PUF domain of which it is a variant - e.g., one that retains the ability to recognize target RNA to a similar extent, the same extent, or to a higher extent in terms of binding affinity, and/or with substantially the same or identical binding specificity, as the parent PUF domain. The functional variant PUF domain can, for instance, be at least 30%, 50%, 75%, 80%, 90%, 98% or more identical in amino acid sequence to the parent PUF domain. The functional variant can, for example, comprise the amino acid sequence of the parent PUF domain with at least one conservative amino acid substitution, for example, conservative amino acid substitutions in the scaffold of the PUF domain (*i.e.*, amino acids that do not interact with the RNA). Alternatively or additionally, the functional variants can comprise the amino acid sequence of the parent PUF domain with at least one non-conservative amino acid substitution. In this case, it is preferable for the non-conservative amino acid substitution to not interfere with or inhibit the biological activity of the functional variant. The non-conservative amino acid substitution may enhance the biological activity of the functional variant, such that the biological activity of the functional variant is increased as compared to the parent PUF domain, or may alter the stability of the PUF domain to a desired level (e.g., due to substitution of amino acids in the scaffold). The PUF domain can consist essentially of the

specified amino acid sequence or sequences described herein, such that other components, *e.g.*, other amino acids, do not materially change the biological activity of the functional variant. In certain some embodiments, the PUF domain is a Pumilio homology domain (PU-HUD). In a particular embodiment, the PU-HUD is a human Pumilio 1 domain. In certain some embodiments, the PUF domain has the sequence of any one of the PUF domains disclosed in international application PCT/US2016/021491, published as WO2016148994 A8, in international application PCT/US2011/040933, published as WO 2011/160052A2, and Spassov & Jurecic ("Cloning and comparative sequence analysis of PUM1 and PUM2 genes, human members of the Pumilio family of RNA-binding proteins," *Gene*, 299:195-204, October 2002), which are hereby incorporated by reference in their entirety and for all purposes.

In some embodiments, the PUF domain includes a PUFa domain, a PUFb domain, a PUFc domain, or a PUFw domain. In certain some embodiments, the PUFa domain has the amino acid sequence of SEQ ID NO: 27.

The subject nucleic acid includes one or more tandem sequences, each of which can be specifically recognized and bound by a specific PUF domain (*infra*). Since a PUF domain can be engineered to bind virtually any PBS sequence based on the nucleotide-specific interaction between the individual PUF motifs of PUF domain and the single RNA nucleotide they recognize, the PBS sequences can be any designed sequence that bind their corresponding PUF domain. Any of the subject PUF domain can be made using, for example, a Golden Gate Assembly kit (see Abil *et al.*, *Journal of Biological Engineering* 8:7, 2014), which is available at Addgene (Kit # 1000000051).

DNA Repair Domains

As described herein the DNA repair domains provided herein are linked to one or more PUF domains, in some embodiments, forming a DNA repair protein. In the presence of the DNA repair domain, the strand break introduced by the RNA-guided endonuclease at a target nucleic acid sequence may be repaired at an increased rate, for example, through HDR, relative to the absence of the DNA repair domain. In some embodiments, the strand is repaired at a decreased rate, for example, through NHEJ relative to the absence of the DNA repair domain. Where the strand break is repaired at a higher rate (*e.g.*, through HDR), the relative amount of strand breaks repaired (*e.g.*, through HDR) is higher than in the absence of the DNA repair domain. In some embodiments, the complexes provided herein are capable of increasing HDR activity at a target site. In some embodiments, the increase is 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%,

100%, or any percent increase in between 10% and 100% as compared to native or control levels.

In some embodiments, HDR activity is increased by about 15% to about 95%. In some embodiments, HDR activity is increased by about 20% to about 95%. In some embodiments, HDR activity is increased by about 25% to about 95%. In some embodiments, HDR activity is increased by about 30% to about 95%. In some embodiments, HDR activity is increased by about 35% to about 95%. In some embodiments, HDR activity is increased by about 40% to about 95%. In some embodiments, HDR activity is increased by about 45% to about 95%. In some embodiments, HDR activity is increased by about 50 to about 95%. In some embodiments, HDR activity is increased by about 55% to about 95%. In some embodiments, HDR activity is increased by about 60% to about 95%. In some embodiments, HDR activity is increased by about 65% to about 95%. In some embodiments, HDR activity is increased by about 70 to about 95%. In some embodiments, HDR activity is increased by about 75% to about 95%. In some embodiments, HDR activity is increased by about 80% to about 95%. In some embodiments, HDR activity is increased by about 85% to about 95%. In some embodiments, HDR activity is increased by about 90% to about 95%.

In some embodiments, the increase is expressed as “-fold” increase. For example, the increase in HDR activity be at least a 1.2-fold, 1.5-fold, 2-fold, 5-fold, or more over a control. In some embodiments, the increase is about 1.2-fold, about 1.3-fold, about 1.4-fold, about 1.5-fold, about 1.6-fold, about 1.7-fold, about 1.8-fold, about 1.9-fold, about 2.0-fold, about 2.2-fold, about 2.3-fold, about 2.4-fold, about 2.5-fold, about 2.6-fold, about 2.7-fold, about 2.8-fold, about 2.9-fold, about 3.0-fold, about 3.2-fold, about 3.3-fold, about 3.4-fold, about 3.5-fold, about 3.6-fold, about 3.7-fold, about 3.8-fold, about 3.9-fold, about 4.0-fold, about 4.2-fold, about 4.3-fold, about 4.4-fold, about 4.5-fold, about 4.6-fold, about 4.7-fold, about 4.8-fold, about 4.9-fold, about 5.0-fold, about 5.2-fold, about 5,3-fold, about 5.4-fold, about 5.5-fold, about 5.6-fold, about 5.7-fold, about 5.8-fold, about 5.9-fold, or about 6.0-fold relative to a control.

A “control” sample or value refers to a sample that serves as a reference, usually a known reference, for comparison to a test sample. For example, a test sample can be a sample including the DNA repair domain provided herein and compared to samples lacking the DNA repair domain, or a known standard sample useful as a negative control. A control value can also be obtained from the same sample, *e.g.*, from an earlier-obtained sample, prior to introducing a

DNA repair complex provided herein. One of skill will recognize that controls can be designed for assessment of any number of parameters.

One of skill in the art will understand which controls are valuable in a given situation and be able to analyze data based on comparisons to control values. Controls are also valuable for determining the significance of data. For example, if values for a given parameter are widely variant in controls, variation in test samples will not be considered as significant.

The term “repair” as provided herein refers to the processes by which a strand break is identified and corrected. The process of strand repair includes several enzymatic steps the completion of which results in the transformation of a strand break into an intact strand. During the process of repair one or more nucleotides may be replaced by new nucleotides thereby changing the sequence composition at and around the site of the strand break. The process of strand break repair includes, for example, ligation, polymerization, endonucleolytic cleavage, and decoiling. Thus, in some embodiments, the DNA repair domain is or comprises a ligase or ligase activity, a polymerase or polymerase activity, a topoisomerase or topoisomerase activity, a helicase or helicase activity, or an endonuclease or endonuclease.

In some embodiments, the DNA repair domain includes an endonuclease domain, a helicase domain or a ligase domain. In some embodiments, the DNA repair domain is an endonuclease domain, a helicase domain or a ligase domain. In some embodiments the DNA repair domain includes a BRCA1 protein or functional fragment thereof, a RAD54L protein or functional fragment thereof, a CtIP protein or functional fragment thereof, a PALB2 protein or functional fragment thereof, a RAD51A protein or functional fragment thereof, a XRCC3 protein or functional fragment thereof, a RECQ5 protein or functional fragment thereof, a FEN1 protein or functional fragment thereof, a FANCB protein or functional fragment thereof, a FANCF protein or functional fragment thereof, a FANCG protein or functional fragment thereof, a FANCM protein or functional fragment thereof, a MRE11A protein or functional fragment thereof, a USP1 protein or functional fragment thereof, a RPA1 protein or functional fragment thereof, a RPA2 protein or functional fragment thereof, a BRC3 protein or functional fragment thereof, or a BRC4 protein or functional fragment thereof.

In some embodiments, the DNA repair domain is a BRCA1 protein or functional fragment thereof, a RAD54L protein or functional fragment thereof, a CtIP protein or functional fragment thereof, a PALB2 protein or functional fragment thereof, a RAD51A protein or functional fragment thereof, a XRCC3 protein or functional fragment thereof, a RECQ5 protein or functional fragment thereof, a FEN1 protein or functional fragment thereof, a FANCB protein

or functional fragment thereof, a FANCF protein or functional fragment thereof, a FANCG protein or functional fragment thereof, a FANCM protein or functional fragment thereof, a MRE11A protein or functional fragment thereof, a USP1 protein or functional fragment thereof, a RPA1 protein or functional fragment thereof, a RPA2 protein or functional fragment thereof, a BRC3 protein or functional fragment thereof, or a BRC4 protein or functional fragment thereof.

The terms "BRCA1," "BRCA1 protein," "BRCA1 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the breast cancer type 1 susceptibility protein (BRCA1) protein or variants or homologs thereof that maintain BRCA1 protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to BRCA1). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring BRCA1 polypeptide. In some embodiments, the BRCA1 peptide is substantially identical to the protein identified by the UniProt reference number P38398 or a variant or homolog having substantial identity thereto. In some embodiments, the BRCA1 peptide is substantially identical to the protein identified by the GenBank reference number AAX42696.1, NP_009225.1 or a variant or homolog having substantial identity thereto.

The terms "RAD54L," "RAD54L protein," "RAD54L peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the DNA repair and recombination protein RAD54-like (RAD54L) protein or variants or homologs thereof that maintain RAD54L protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to RAD54L). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring RAD54L polypeptide. In some embodiments, the RAD54L peptide is substantially identical to the protein identified by the UniProt reference number Q92698 or a variant or homolog having substantial identity thereto. In some embodiments, the RAD54L peptide is substantially identical to the protein identified by the GenBank reference number CAA66379.1 or a variant or homolog having substantial identity thereto.

The terms "CtIP," "CtIP protein," "CtIP peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the C-terminal binding protein Interacting Protein (CtIP) protein or variants or homologs thereof that maintain CtIP protein activity (e.g. within at

least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to CtIP). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring CtIP polypeptide. In some embodiments, the CtIP peptide is substantially identical to the protein identified by the UniProt reference number Q99708 or a variant or homolog having substantial identity thereto. In some embodiments, the CtIP peptide is substantially identical to the protein identified by the GenBank reference number NP_002885.1 or a variant or homolog having substantial identity thereto.

The terms "PALB2," "PALB2 protein," "PALB2 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the Partner and Localizer of BRCA2 (PALB2) protein or variants or homologs thereof that maintain PALB2 protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to PALB2). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring PALB2 polypeptide. In some embodiments, the PALB2 peptide is substantially identical to the protein identified by the UniProt reference number Q86YC2 or a variant or homolog having substantial identity thereto. In some embodiments, the PALB2 peptide is substantially identical to the protein identified by the GenBank reference number NP_078951.2 or a variant or homolog having substantial identity thereto.

The terms "RAD51A," "RAD51A protein," "RAD51A peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the DNA repair protein RAD51 homolog 1 (RAD51A) protein or variants or homologs thereof that maintain RAD51A protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to RAD51A). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring RAD51A polypeptide. In some embodiments, the RAD51A peptide is substantially identical to the protein identified by the UniProt reference number Q06609 or a variant or homolog having substantial identity thereto. In some embodiments, the RAD51A peptide is substantially identical to the protein identified by the GenBank reference number NP_002866.2 or a variant or homolog having substantial identity thereto.

The terms "XRCC3," "XRCC3 protein," "XRCC3 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the X-ray Repair Cross Complementing 3 (XRCC3) protein or variants or homologs thereof that maintain XRCC3 protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to XRCC3). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring XRCC3 polypeptide. In some embodiments, the XRCC3 peptide is substantially identical to the protein identified by the UniProt reference number O43542 or a variant or homolog having substantial identity thereto. In some embodiments, the XRCC3 peptide is substantially identical to the protein identified by the GenBank reference number NP_005423.1 or a variant or homolog having substantial identity thereto.

The terms "RECQ5," "RECQ5 protein," "RECQ5 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the ATP-dependent DNA Helicase Q5 (RECQ5) protein or variants or homologs thereof that maintain RECQ5 protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to RECQ5). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring RECQ5 polypeptide. In some embodiments, the RECQ5 peptide is substantially identical to the protein identified by the UniProt reference number O94762 or a variant or homolog having substantial identity thereto. In some embodiments, the RECQ5 peptide is substantially identical to the protein identified by the GenBank reference number NP_005423.1 or a variant or homolog having substantial identity thereto.

The terms "FEN1," "FEN1 protein," "FEN1 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the Flap endonuclease 1 (FEN1) protein or variants or homologs thereof that maintain FEN1 protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to FEN1). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring FEN1 polypeptide. In some embodiments, the FEN1 peptide is substantially identical to the protein identified by the UniProt reference number P39748 or a variant or homolog having substantial identity thereto. In

some embodiments, the FEN1 peptide is substantially identical to the protein identified by the GenBank reference number NP_004102.1 or a variant or homolog having substantial identity thereto.

The terms "FANCB," "FANCB protein," "FANCB peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the Fanconi anemia group B (FANCB) protein or variants or homologs thereof that maintain FANCB protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to FANCB). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring FANCB polypeptide. In some embodiments, the FANCB peptide is substantially identical to the protein identified by the UniProt reference number Q8NB91 or a variant or homolog having substantial identity thereto. In some embodiments, the FANCB peptide is substantially identical to the protein identified by the GenBank reference number NP_689846.1 or a variant or homolog having substantial identity thereto.

The terms "FANCF," "FANCF protein," "FANCF peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the Fanconi anemia group F (FANCF) protein or variants or homologs thereof that maintain FANCF protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to FANCF). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring FANCF polypeptide. In some embodiments, the FANCF peptide is substantially identical to the protein identified by the UniProt reference number Q9NPI8 or a variant or homolog having substantial identity thereto. In some embodiments, the FANCF peptide is substantially identical to the protein identified by the GenBank reference number NP_073562.1 or a variant or homolog having substantial identity thereto.

The terms "FANCG," "FANCG protein," "FANCG peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the Fanconi anemia group G (FANCG) protein or variants or homologs thereof that maintain FANCG protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to FANCG). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10,

20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring FANCG polypeptide. In some embodiments, the FANCG peptide is substantially identical to the protein identified by the UniProt reference number O15287 or a variant or homolog having substantial identity thereto. In some embodiments, the FANCG peptide is substantially identical to the protein identified by the GenBank reference number NP_004620.1 or a variant or homolog having substantial identity thereto.

The terms "FANCM," "FANCM protein," "FANCM peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the Fanconi anemia group M (FANCM) protein or variants or homologs thereof that maintain FANCM protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to FANCM). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring FANCM polypeptide. In some embodiments, the FANCM peptide is substantially identical to the protein identified by the UniProt reference number Q8IYD8 or a variant or homolog having substantial identity thereto. In some embodiments, the FANCM peptide is substantially identical to the protein identified by the GenBank reference number NP_065988.1 or a variant or homolog having substantial identity thereto.

The terms "MRE11A," "MRE11A protein," "MRE11A peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the double-strand break repair protein (MRE11A) or variants or homologs thereof that maintain MRE11A protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to MRE11A). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring MRE11A polypeptide. In some embodiments, the MRE11A peptide is substantially identical to the protein identified by the UniProt reference number P49959 or a variant or homolog having substantial identity thereto. In some embodiments, the MRE11A peptide is substantially identical to the protein identified by the GenBank reference number NP_005582.1 or a variant or homolog having substantial identity thereto.

The terms "USP1," "USP1 protein," "USP1 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the ubiquitin specific peptidase 1 (USP1) or variants or homologs thereof that maintain USP1 protein activity (e.g. within at least 50%, 80%,

90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to USP1). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring USP1 polypeptide. In some embodiments, the USP1 peptide is substantially identical to the protein identified by the UniProt reference number O94782 or a variant or homolog having substantial identity thereto. In some embodiments, the USP1 peptide is substantially identical to the protein identified by the GenBank reference number NP_003359.3 or a variant or homolog having substantial identity thereto.

The terms " RPA1," "RPA1 protein," "RPA1 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the replication protein A1 (RPA1) protein or variants or homologs thereof that maintain RPA1 protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to RPA1). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring RPA1 polypeptide. In some embodiments, the RPA1 peptide is substantially identical to the protein identified by the UniProt reference number P27694 or a variant or homolog having substantial identity thereto. In some embodiments, the RPA1 peptide is substantially identical to the protein identified by the GenBank reference number NP_002936.1 or a variant or homolog having substantial identity thereto.

The terms " RPA2," "RPA2 protein," "RPA2 peptide" as referred to herein include any of the recombinant or naturally-occurring forms of the replication protein A2 (RPA2) protein or variants or homologs thereof that maintain RPA2 protein activity (e.g. within at least 50%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% activity compared to RPA2). In some aspects, the variants or homologs have at least 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid sequence identity across the whole sequence or a portion of the sequence (e.g. a 10, 20, 50, 100, 150 or 200 continuous amino acid portion) compared to a naturally occurring RPA2 polypeptide. In some embodiments, the RPA2 peptide is substantially identical to the protein identified by the UniProt reference number P15927 or a variant or homolog having substantial identity thereto. In some embodiments, the RPA2 peptide is substantially identical to the protein identified by the GenBank reference number NP_002937.1 or a variant or homolog having substantial identity thereto.

In some embodiments, the DNA repair domain is linked to the C-terminus of the PUF domain. In some embodiments, the DNA repair domain is linked to the N-terminus of the PUF domain.

In some embodiments, the DNA repair protein (e.g., conjugate) includes the sequence of SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, or SEQ ID NO: 80. In some embodiments, the DNA repair protein has the sequence of SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, or SEQ ID NO: 80. The DNA repair protein as provided herein may further include a nuclear localization sequence (NLS).

In some embodiments, the DNA repair protein (e.g., conjugate) includes in N-terminal to C-terminal direction: a DNA repair domain, a first peptide linker including a first, a second, a third and a fourth nuclear localization sequence, a PUFa domain, a second peptide linker including a fifth and a sixth nuclear localization sequence.

In some embodiments, the DNA repair complex further includes a donor nucleic acid including a donor sequence. In some embodiments, the donor nucleic acid is single stranded or double-stranded. In some embodiments, the donor nucleic acid forms part of a circular DNA molecule (e.g., plasmid, vector). In some embodiments, the donor nucleic acid forms part of a linear DNA molecule (e.g., oligonucleotide). In some embodiments, the donor nucleic acid has the sequence of SEQ ID NO: 85.

Additional Programmable Nucleases

Some aspects of the present disclosure provide DNA repair proteins comprising a programmable nuclease linked to DNA repair domain selected from the group consisting of: RPA1; RPA2; FANCM; BRCA1; RAD54L; PALB2; XRCC3; FEN1; RECQ5; FANCB; USP1; FANCF; and FANCG. In some embodiments, the programmable nuclease comprises a RNA-guided nuclease. For example, the RNA-guided nuclease may be Cas9 nuclease or Cas9 nickase.

In some embodiments, the programmable nuclease comprises a ZFN. In other embodiments, the programmable nuclease comprises a TALEN.

Zinc-Finger Nucleases

Methods described herein, in some embodiments, include the use of zinc-finger nucleases. A zinc-finger nuclease (ZFN) is an endonuclease that can be programmed to cut specific sequences of DNA. ZFNs are composed of a zinc-finger DNA-binding domain and a nuclease domain.

The DNA-binding domains of individual ZFNs generally contain 3-6 individual zinc finger repeats that recognize 9-18 nucleotides. For example, if the zinc finger domain perfectly recognizes a 3 base pair sequence, then a 3 zinc finger array can be generated to recognize a 9 base pair target DNA sequence. Because individual zinc fingers recognize relatively short (*e.g.*, 3 base pairs) target DNA sequences, ZFNs with 4, 5, or 6 zinc finger domains are typically used to minimize off-target DNA cutting. Non-limiting examples of zinc finger DNA-binding domains that may be used with methods of the present disclosure include Zif268, Gal4, HIV nucleocapsid protein, MYST family histone acetyltransferases, myelin transcription factor Myt1, and suppressor of tumorigenicity protein 18 (ST18). A ZFN may contain homogeneous DNA binding domains (all from the same source molecule) or a ZFN may contain heterogeneous DNA binding domains (at least one DNA binding domain is from a different source molecule).

Zinc finger DNA-binding domains work in concert with a nuclease domain to form a zinc finger nucleases (ZFNs) that cut target sequences. The nuclease cuts the DNA in a non-sequence specific manner after being recruited to a target sequence by the zinc fingers DNA-binding domains. The most widely-used ZFN is the type II restriction enzyme FokI, which forms a heterodimer before producing a double-stranded break in the DNA. Thus, two ZFN proteins must bind to opposite strands of DNA to create the FokI heterodimer and form a double-stranded break, reducing off-target DNA cleavage events (Kim, et al., *Proc Natl Acad Sci USA*, 1996, 93(3): 1156-1160). Additionally, ZFNs may be nickases that only cleave one strand of the double-stranded DNA. By cleaving only one strand, the DNA is more likely to be repaired by error-free HR as opposed to error-prone NHEJ (Ramirez, et al., *Nucleic Acids Research*, 40(7): 5560-5568). Non-limiting examples of nucleases that may be used with methods in this disclosure include FokI and DNaseI.

It should be understood that the ZFN in the ZFN-based gene editing system may be expressed as a fusion protein, with the DNA-binding domain and the nuclease domain expressed in the same polypeptide. This fusion may include a linker of amino acids (*e.g.*, 1, 2, 3, 4, 5, 6, or more) between the DNA-binding domain and the nuclease domain.

Transcription Activator-Like Effector Nucleases

Methods described herein, in some embodiments, include the use of transcription activator-like effector nucleases (TALENs). A TALEN is an endonuclease that can be programmed to cut specific sequences of DNA. TALENs are composed of transcription activator-like effector (TALE) DNA-binding domains, which recognize single target nucleotides in the DNA, and transcription activator-like effector nucleases (TALENs) which cut the DNA at or near a target nucleotide.

Transcription activator-like effectors (TALEs) found in bacteria are modular DNA binding domains that include central repeat domains made up of repetitive sequences of residues (Boch J. et al. *Annual Review of Phytopathology* 2010; 48: 419–36; Boch J *Biotechnology* 2011; 29(2): 135-136). The central repeat domains, in some embodiments, contain between 1.5 and 33.5 repeat regions, and each repeat region may be made of 34 amino acids; amino acids 12 and 13 of the repeat region, in some embodiments, determines the nucleotide specificity of the TALE and are known as the repeat variable diresidue (RVD) (Moscou MJ et al. *Science* 2009; 326 (5959): 1501; Juillerat A et al. *Scientific Reports* 2015; 5: 8150). Unlike ZF DNA sensors, TALE-based sequence detectors can recognize single nucleotides. In some embodiments, combining multiple repeat regions produces sequence-specific synthetic TALEs (Cermak T et al. *Nucleic Acids Research* 2011; 39 (12): e82). Non-limiting examples of TALEs that may be utilized in the present disclosure include IL2RG, AvrBs, dHax3, and thXoI

A transcription activator-like effector nuclease (TALEN) cleaves the DNA non-specifically after being recruited to a target sequence by the TALE. This non-specific cleavage can lead to off-target DNA cleavage events. The most widely-used TALEN is the type II restriction enzyme FokI, which forms a heterodimer to produce a double-stranded break in DNA. Thus, two TALEN proteins must bind to opposite strands of DNA to create the FokI heterodimer and form a double-stranded break, reducing off-target DNA cleavage events (Christian M et al. *Genetics* 2010; 186: 757-761). Additionally, TALEN nucleases may be nickases, which cut only a single-strand of the DNA, thus promoting repair of the break by HR (Gabsalilow L. et al. *Nucleic Acids Res.* 41, e83). Non-limiting examples of TALENs that may be utilized in the present disclosure include FokI, RNaseH, and MutH.

It should be understood that the TALEN in the TALEN-based gene editing system may be expressed as a fusion protein, with the DNA-binding domain and the nuclease domain expressed in the same polypeptide. This fusion may include a linker of amino acids (*e.g.*, 1, 2, 3, 4, 5, 6, or more) between the DNA-binding domain and the nuclease domain.

Donor Nucleic Acids

A donor nucleic acid is a nucleic acid that includes a sequence of interest. In some embodiments, a donor nucleic acid comprise a sequence that is partially complementary to a target nucleic acid sequence. In some embodiments, a donor nucleic acid comprise a sequence that is homologous to a target nucleic acid sequence. A donor nucleic acid, in some embodiments, includes one or more continuous nucleotide sequences that are homologous to a target nucleic acid sequence (e.g., a left and a right homology arm) and includes a nucleotide sequence that is non homologous to a target nucleic acid sequence (also referred to herein as insert). Thus, in some embodiments, a donor nucleic acid includes one or more homologous donor sequences. In some embodiments, a donor nucleic acid includes a first homologous donor sequence and a second homologous donor sequence, wherein the first and the second homologous donor sequence are connected through a non-homologous insert sequence.

In some embodiments, a donor sequence is about 5 to about 2,500, about 5 to about 2000, about 5 to about 1500, about 5 to about 1000, about 5 to about 500, about 5 to about 250, about 5 to about 100, or about 5 to about 50 nucleotides in length. In some embodiments, a donor sequence is about 10 to about 2,500, about 10 to about 2000, about 10 to about 1500, about 10 to about 1000, about 10 to about 500, about 10 to about 250, about 10 to about 100, or about 10 to about 50 nucleotides in length. In some embodiments, a donor sequence is about 20 to about 2,500, about 20 to about 2000, about 20 to about 1500, about 20 to about 1000, about 20 to about 500, about 20 to about 250, about 20 to about 100, or about 20 to about 50 nucleotides in length. In some embodiments, a donor sequence is about 5, 10, 15, 20, or 25 nucleotides in length. In some embodiments, a donor sequence is about 50 nucleotides in length. In some embodiments, a donor sequence is about 100 nucleotides in length. In some embodiments, a donor sequence is about 150 nucleotides in length. In some embodiments, a donor sequence is about 200 nucleotides in length. In some embodiments, a donor sequence is about 500 nucleotides in length. In some embodiments, a donor sequence is about 1,000 nucleotides in length. In some embodiments, a donor sequence is about 2,000 nucleotides in length.

In some embodiments, a first or second homologous donor sequence are independently from about 5 to about 2,500 nucleotides in length. In some embodiments, a first and second homologous donor sequence are independently about 5, 10, 15, 20, or 25 nucleotides in length. In some embodiments, a first and second homologous donor sequence are independently about 50 nucleotides in length. In some embodiments, the first and second homologous donor sequence are independently about 100 nucleotides in length. In some embodiments, a first and second

homologous donor sequence are independently about 150 nucleotides in length. In some embodiments, a first and second homologous donor sequence are independently about 200 nucleotides in length. In some embodiments, a first and second homologous donor sequence are independently about 500 nucleotides in length. In some embodiments, a first and second homologous donor sequence are independently about 1,000 nucleotides in length. In some embodiments, a first and second homologous donor sequence are independently about 2,000 nucleotides in length.

In some embodiments, the first and second homologous donor sequence have the same or different nucleotide lengths. In some embodiments, the first homologous donor sequence is 36 nucleotides in length. In some embodiments, the second homologous donor sequence is 91 nucleotides in length. In some embodiments, the first homologous donor sequence is 36 nucleotides in length and the second homologous donor sequence is 91 nucleotides in length. Any of the donor nucleic acids described in Richardson et al. (Nat Biotechnol. 2016 Mar; 34 (3):339-44), which is hereby incorporated in its entirety and for all purposes, may be used for the compositions and methods provided herein including embodiments thereof.

Cells

In another aspect, a cell comprising a DNA repair complex as provided herein including embodiments thereof is provided. In some embodiments, the cell is a mammalian cell.

Exemplary cell types contemplated as described herein, include, a bacterial cell; an archaeal cell; a single-celled eukaryotic organism; a plant cell; an algal cell, *e.g.*, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh*, and the like; a fungal cell; an animal cell; a cell from an invertebrate animal (*e.g.*, an insect, a cnidarian, an echinoderm, a nematode, *etc.*); a eukaryotic parasite (*e.g.*, a malarial parasite, *e.g.*, *Plasmodium falciparum*; a helminth; *etc.*); a cell from a vertebrate animal (*e.g.*, fish, amphibian, reptile, bird, mammal); a mammalian cell, *e.g.*, a rodent cell, a human cell, a non-human primate cell. Suitable host cells include naturally occurring cells; genetically modified cells (*e.g.*, cells genetically modified in a laboratory, *e.g.*, by the “hand of man”); and cells manipulated *in vitro* in any way. In some embodiments, a host cell is isolated or cultured.

Any type of cell may be of interest (*e.g.*, a stem cell, *e.g.* an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell, a germ cell; a somatic cell, *e.g.* a fibroblast, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell; an *in vitro* or *in vivo*

embryonic cell of an embryo at any stage, *e.g.*, a 1-cell, 2-cell, 4-cell, 8-cell, *etc.* stage zebrafish embryo; *etc.*). Cells may be from established cell lines or they may be primary cells, where “primary cells,” “primary cell lines,” and “primary cultures” are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow *in vitro* for a limited number of passages, *i.e.* splittings, of the culture. For example, primary cultures include cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times go through the crisis stage. Primary cell lines can be maintained for fewer than 10 passages *in vitro*. Target cells are in many embodiments, unicellular organisms, or are grown in culture.

If the cells are primary cells, such cells may be harvest from an individual by any convenient method. For example, leukocytes may be conveniently harvested by apheresis, leukocytapheresis, density gradient separation, *etc.*, while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, *etc.* are most conveniently harvested by biopsy. An appropriate solution may be used for dispersion or suspension of the harvested cells. Such solution will generally be a balanced salt solution, *e.g.* normal saline, phosphate-buffered saline (PBS), Hank’s balanced salt solution, *etc.*, conveniently supplemented with fetal calf serum or other naturally occurring factors, in conjunction with an acceptable buffer at low concentration, *e.g.*, from 5-25 mM. Convenient buffers include HEPES, phosphate buffers, lactate buffers, *etc.* The cells may be used immediately, or they may be stored, frozen, for long periods of time, being thawed and capable of being reused. In such cases, the cells will usually be frozen in 10% dimethyl sulfoxide (DMSO), 50% serum, 40% buffered medium, or other solutions commonly used in the art to preserve cells at such freezing temperatures, and thawed in any suitable manner.

In some embodiments, the cell is a cancer cell.

Another aspect of the present disclosure provides a host cell including any one of the subject vector, nucleic acid, and complex.

In some embodiments, the RNA-guided DNA endonuclease is encoded by a first nucleic acid. In some embodiments, the nucleic acid (*i.e.*, guide RNA) is encoded by a second nucleic acid. In some embodiments, the DNA repair protein (*e.g.*, conjugate) is encoded by a third nucleic acid. In some embodiments, the donor sequence is encoded by a fourth nucleic acid. In some embodiments, the expression of the RNA-guided DNA endonuclease, the nucleic acid (*i.e.*, guide RNA), the DNA repair protein, or the donor sequence can be under the control of a

constitutive promoter or an inducible promoter. In some embodiments, the cell includes the first nucleic acid, second nucleic acid, the third nucleic acid or the fourth nucleic acid.

In some embodiments, the first nucleic acid is contained within a first vector. In some embodiments, the second nucleic acid is contained within a second vector. In some embodiments, the third nucleic acid is contained within a third vector. In some embodiments, the fourth nucleic acid is contained within a fourth vector. In some embodiments, either the first, second, third or fourth vector is the same. In some embodiments, the first, second, third or fourth vector is a transfection vector. In some embodiments, the first, second, third or fourth vector is a viral vector. In some embodiments, the cell includes the first, second, third or fourth vector. In some embodiments, the cell includes the first, second, third and fourth vector.

In some embodiments, sequences that can be encoded by different vectors may be on the same vector. For example, in some embodiments, the second vector may be the same as the vector, and/or the third vector may be the same as the vector or the second vector.

The host cell may be in a live animal, or may be a cultured cell.

Nucleic Acid Expression

In some embodiments, a RNA-guided DNA endonuclease (e.g., Cas9) is encoded by a first nucleic acid. In some embodiments, the nucleic acid (i.e., guide RNA) is encoded by a second nucleic acid. In some embodiments, the DNA repair protein (e.g., conjugate) is encoded by a third nucleic acid. In some embodiments, the donor sequence is encoded by a fourth nucleic acid.

In some embodiments, the first nucleic acid is contained within a first vector. In some embodiments, the second nucleic acid (i.e., guide RNA) is contained within a second vector. In some embodiments, the third nucleic acid is contained within a third vector. In some embodiments, the fourth nucleic acid is contained within a fourth vector. In some embodiments, either the first, second, third or fourth vector is the same. In some embodiments, the first, second, third or fourth vector is a transfection vector. In some embodiments, the first, second, third or fourth vector is a viral vector.

A nucleic acid, a nucleic acid comprising a nucleotide sequence encoding same, or a nucleic acid comprising a nucleotide sequence encoding the subject RNA-guided DNA endonuclease (Cas9 protein), nucleic acid (i.e., guide RNA) or DNA repair protein (e.g., conjugate), can be introduced into a host cell by any of a variety of methods.

Any method can be used to introduce a nucleic acid (*e.g.*, vector or expression construct) into a stem cell or progenitor cell. Examples of these methods include, include viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, *e.g.*, Panyam *et al.*, *Adv. Drug Deliv. Rev.*, pii: S0169-409X(12)00283-9.doi:10.1016 / j.addr.2012.09.023), and the like.

In some embodiments, a method involves introducing into a host cell (or a population of host cells) one or more nucleic acids (*e.g.*, vectors) comprising nucleotide sequences encoding a subject nucleic acid and/or a RNA-guided DNA endonuclease (Cas9 protein) and/or a DNA repair protein (*e.g.*, conjugate). In some embodiments, a host cell including a target DNA is *in vitro*. In some embodiments, a host cell including a target DNA is *in vivo*. Suitable nucleic acids including nucleotide sequences encoding a subject nucleic acid and/or a RNA-guided DNA endonuclease (Cas9 protein) and/or a DNA repair protein include expression vectors, where the expression vectors may be recombinant expression vector.

In some embodiments, the recombinant expression vector is a viral construct, *e.g.*, a recombinant adeno-associated virus construct (see, *e.g.*, U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, a recombinant retroviral construct, *etc.*

Suitable expression vectors include, but are not limited to, viral vectors (*e.g.* viral vectors based on vaccinia virus; poliovirus; adenovirus (see, *e.g.*, Li *et al.*, *Invest Ophthalmol. Vis. Sci.*, 35:2543-2549, 1994; Borrás *et al.*, *Gene Ther.*, 6:515-524, 1999; Li and Davidson, *Proc. Natl. Acad. Sci. USA*, 92:7700-7704, 1995; Sakamoto *et al.*, *Hum. Gene Ther.*, 5:1088-1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, *e.g.*, Ali *et al.*, *Hum. Gene Ther.*, 9:81-86, 1998, Flannery *et al.*, *Proc. Natl. Acad. Sci. USA*, 94:6916-6921, 1997; Bennett *et al.*, *Invest Ophthalmol Vis Sci* 38:2857-2863, 1997; Jomary *et al.*, *Gene Ther.*, 4:683-690, 1997, Rolling *et al.*, *Hum. Gene Ther.*, 10:641-648, 1999; Ali *et al.*, *Hum. Mol. Genet.*, 5:591-594, 1996; Srivastava in WO 93/09239, Samulski *et al.*, *J. Vir.*, 63:3822-3828, 1989; Mendelson *et al.*, *Viol.*, 166: 154-165, 1988; and Flotte *et al.*, *Proc. Natl. Acad. Sci. USA*, 90: 10613-10617, 1993); SV40; herpes simplex virus; human immunodeficiency virus (see, *e.g.*, Miyoshi *et al.*, *Proc. Natl. Acad. Sci. USA*, 94: 10319-23, 1997; Takahashi *et al.*, *J. Virol.*, 73:7812-7816, 1999); a retroviral vector

(e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, HIV virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

Numerous suitable expression vectors may be used, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell. Any one of the vectors described in international application PCT/US2016/021491 and published as WO2016148994 A8, which is hereby incorporated by reference and for all purposes, is contemplated for the complexes and methods provided herein including embodiments thereof.

Methods of Modulating DNA repair

The methods and complexes provided herein provide, *inter alia*, for a versatile delivery platform of DNA repair activities, which promote HDR activities. Using the methods and complexes provided herein DNA repair domains (e.g., BRCA1, FEN1 enzymes or functional fragments thereof) or combinations thereof, may be delivered to a cell sequentially or concomitantly. Delivery of a combination of DNA repair domains to a cell, allows for an increase of HDR activity at a targeted gene locus. The present disclosure further provides for the delivery of a plurality of DNA repair domains, wherein the domains may be the same or different. Where a plurality of DNA repair domains is delivered to a cell, the domains may form part of a plurality of DNA repair proteins (e.g., conjugates), each linked to a PUF domain, and/or they may be directly fused to the RNA-guided DNA endonuclease enzyme (e.g., Cas9). Further, and by virtue of a target-gene specificity of the guide RNA, the present disclosure allows for the delivery of DNA repair domains to different target sites in a cell at the same time. Applicants were the first to show that due to targeted delivery of DNA repair domain to a, for example, Cas9-introduced strand break, the activity of HDR at the break can be increased relative to the absence of the DNA repair domain.

In another aspect, a method of increasing homology directed repair (HDR) in a mammalian cell is provided. The method includes: (a) providing a mammalian cell containing a target nucleic acid requiring homology directed repair; (b) delivering to the mammalian cell a first nucleic acid encoding an RNA-guided DNA endonuclease; (c) delivering to the mammalian cell a second nucleic acid including: (i) a DNA-targeting sequence that is complementary to a target nucleic acid sequence; (ii) a binding sequence for the RNA-guided DNA endonuclease

enzyme; and (iii) one or more PUF binding site (PBS) sequences, wherein the RNA-guided DNA endonuclease enzyme is capable of binding to the second nucleic acid via the binding sequence; (d) delivering to the mammalian cell a third nucleic acid including a DNA repair protein (e.g., conjugate) including: (i) a PUF domain; and (ii) a DNA repair domain, the DNA repair domain linked to the PUF domain; and (e) delivering to the mammalian cell a fourth nucleic acid encoding a donor sequence; wherein the delivered DNA repair protein increases homology directed repair at a target nucleic acid sequence in the cell relative to a control.

In some embodiments, the DNA repair protein (e.g., conjugate) is bound to the second nucleic acid via binding of the PUF domain to the one or more PBS sequences. In some embodiments, the first nucleic acid is contained within a first vector. In some embodiments, the third nucleic acid is contained within a third vector. In some embodiments, the fourth nucleic acid is contained within a fourth vector. In some embodiments, the first, second, third or fourth vector are the same. In some embodiments, the delivering is performed by transfection.

In some embodiments, the delivered DNA repair protein (e.g., conjugate) is capable of decreasing non-homologous end joining (NHEJ) at a target nucleic acid sequence in the cell relative to a standard control.

In another aspect, a method of decreasing non-homologous end joining (NHEJ) in a mammalian cell is provided. The method includes: (a) providing a mammalian cell containing a target nucleic acid requiring NHEJ; (b) delivering to the mammalian cell a first nucleic acid encoding an RNA-guided DNA endonuclease; (c) delivering to the mammalian cell a second nucleic acid including: (i) a DNA-targeting sequence that is complementary to a target nucleic acid sequence; (ii) a binding sequence for the RNA-guided DNA endonuclease; and (iii) one or more PUF binding site (PBS) sequences, wherein the RNA-guided DNA endonuclease is capable of binding to the second nucleic acid via the binding sequence; (d) delivering to the mammalian cell a third nucleic acid encoding a DNA repair protein (e.g., conjugate) including: (i) a PUF domain; and (ii) a DNA repair domain, the DNA repair domain linked to the PUF domain; and (e) delivering to the mammalian cell a fourth nucleic acid including a donor sequence; wherein the delivered DNA repair protein decreases non-homologous end joining (NHEJ) at a target nucleic acid sequence in the cell relative to a control.

In some embodiments, the DNA repair protein is bound to the second nucleic acid via binding of the PUF domain to the one or more PBS sequences.

In some embodiments, the first nucleic acid is contained within a first vector. In some embodiments, the second nucleic acid is contained within a second vector. In some

embodiments, the third nucleic acid is contained within a third vector. In some embodiments, the fourth nucleic acid is contained within a fourth vector. In some embodiments, the first, second, third or fourth vector are the same. In some embodiments, the delivering is performed by transfection. In some embodiments, the delivered DNA repair protein is capable of increasing HDR at a target nucleic acid sequence in the cell relative to a standard control.

Kits

In another aspect, a kit is provided. The kit includes: (i) a ribonucleoprotein complex as provided herein including embodiments thereof or a nucleic acid encoding the same; and (ii) a DNA repair protein conjugate as provided herein including embodiments thereof or a nucleic acid encoding the same.

In an aspect, the kit includes an RNA-guided DNA endonuclease and a DNA repair protein (e.g., conjugate). The RNA-guided DNA endonuclease and the DNA repair protein may be any of the RNA-guided DNA endonucleases and DNA repair protein provided herein including embodiments thereof (e.g., a Cas9 nickase and a DNA repair protein including a OUF domain linked to a ligase domain). In some embodiments, the kit includes a nucleic acid (e.g., a first nucleic acid) encoding the RNA-guided DNA endonuclease and a nucleic acid (e.g., a third nucleic acid) encoding the DNA repair protein.

In another aspect, the kit includes an RNA-guided DNA endonuclease, a DNA repair protein, a nucleic acid (i.e., guide RNA) and a donor sequence. The RNA-guided DNA endonuclease, the DNA repair protein, the nucleic acid (i.e., guide RNA) and the donor sequence may be any of the RNA-guided DNA endonucleases, DNA repair proteins, nucleic acids (i.e., guide RNAs) and donor sequences provided herein including embodiments thereof. In some embodiments, the kit includes a nucleic acid (e.g., a first nucleic acid) encoding the RNA-guided DNA endonuclease, a nucleic acid (e.g., a third nucleic acid) encoding the DNA repair protein, a nucleic acid (e.g., a second nucleic acid) encoding the nucleic acid (i.e., guide RNA) and a nucleic acid (e.g., a fourth nucleic acid) encoding the donor sequence.

In some embodiments, the kit includes a first nucleic acid encoding the RNA-guided DNA endonuclease, a second nucleic acid encoding the nucleic acid (i.e., guide RNA), a third nucleic acid encoding the DNA repair protein or a fourth nucleic acid encoding the donor sequence. The kit includes a first nucleic acid encoding the RNA-guided DNA endonuclease, a second nucleic acid encoding the nucleic acid (i.e., guide RNA), a third nucleic acid encoding the DNA repair protein and a fourth nucleic acid encoding the donor sequence. In some

embodiments, the first, second, third or fourth nucleic acid independently a form part of a vector. In some embodiments, the first, second, third or fourth vector is a transfection vector. The kits provided herein including embodiments thereof may include nucleic acids (DNA or RNA nucleic acids) encoding the individual components (i.e., RNA-guided DNA endonuclease, the nucleic acid (i.e., guide RNA), the DNA repair protein and the donor sequence), they may include the RNA-guided DNA endonuclease and/or the DNA repair protein as proteins or any combination thereof. In some embodiments, the nucleic acid (i.e., guide RNA) is a guide RNA.

In some embodiments, the kit further includes a transfection agent. In some embodiments, the kit further includes a sample collection device for collecting a sample from a patient.

In some embodiments, a subject kit may include: a) a nucleic acid of the present disclosure, or a nucleic acid (e.g., vector) including a nucleotide sequence encoding the same; optionally, b) a subject Cas9 protein (e.g., Cas9 endonuclease or Cas9 nickase), or a vector encoding the same (including an expressible mRNA encoding the same); and optionally, c) one or more subject DNA repair proteins each including a PUF domain linked to a DNA repair domain that may be the same or different among the different DNA repair proteins, or a vector encoding the same (including an expressible mRNA encoding the same).

In some embodiments, one or more of a) - c) may be encoded by the same vector.

In some embodiments, the kit also comprises one or more buffers or reagents that facilitate the introduction of any one of a) - c) into a host cell, such as reagents for transformation, transfection, or infection.

For example, a subject kit can further include one or more additional reagents, where such additional reagents can be selected from: a buffer; a wash buffer; a control reagent; a control expression vector or RNA nucleic acid; a reagent for *in vitro* production of the Cas9 endonuclease or Cas9 nickase from DNA; and the like.

Components of a subject kit can be in separate containers; or can be combined in a single container.

In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, *etc.* As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or sub-packaging) *etc.* In other

embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, *e.g.* CD-ROM, diskette, flash drive, *etc.* In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, *e.g.* via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

The present disclosure encompasses the following additional embodiments:

1. A DNA repair complex, comprising:
 - (a) a ribonucleoprotein complex comprising:
 - (i) an RNA-guided DNA endonuclease; and
 - (ii) a polynucleotide comprising:
 - (1) a DNA-targeting sequence that is complementary to a target polynucleotide sequence;
 - (2) a binding sequence for the RNA-guided DNA endonuclease; and
 - (3) one or more PUF binding site (PBS) sequences,wherein the RNA-guided DNA endonuclease is bound to the polynucleotide via the binding sequence; and
 - (b) a DNA repair protein conjugate comprising:
 - (i) a PUF domain, the PUF domain having a C-terminus and an N-terminus; and
 - (ii) a DNA repair domain, the DNA repair domain operably linked to the PUF domain to form a DNA repair protein conjugate,wherein the DNA repair protein conjugate binds to the ribonucleoprotein complex via the PUF domain binding to the one or more PBS sequences to form a DNA repair complex, and wherein when the RNA-guided DNA endonuclease introduces a strand break at the target polynucleotide sequence present in a genome, the DNA repair protein conjugate repairs the strand break favoring homology-directed repair (HDR).
2. The DNA repair complex of embodiment 1, further comprising a donor polynucleotide comprising a donor sequence complementary to the DNA-targeting sequence.
3. The DNA repair complex of embodiment 2, wherein the donor polynucleotide is single stranded or double-stranded.

4. The DNA repair complex of any one of embodiments 1-3, wherein the DNA repair domain is operably linked to the C-terminus of the PUF domain to form a DNA repair protein conjugate.
5. The DNA repair complex of any one of embodiments 1-3, wherein the DNA repair domain is operably linked to the N-terminus of the PUF domain to form a DNA repair protein conjugate.
6. The DNA repair complex of one of embodiments 1-5, wherein the RNA-guided DNA endonuclease is a Cas9 nuclease.
7. The DNA repair complex of one of embodiments 1-5, wherein the RNA-guided DNA endonuclease is a Cas9 nickase.
8. The DNA repair complex of one of embodiments 1-7, wherein the DNA repair domain is an endonuclease domain, a helicase domain or a ligase domain.
9. The DNA repair complex of one of embodiments 1-8, wherein the target polynucleotide sequence is part of a gene.
10. The DNA repair complex of one of embodiments 1-8, wherein the target polynucleotide sequence is part of a transcriptional regulatory sequence.
11. The DNA repair complex of one of embodiments 1-10, wherein the one or more PBS sequences contain 8 nucleotides in length.
12. The DNA repair complex of one of embodiments 1-11, wherein the one or more PBS sequences are identical.
13. The DNA repair complex of one of embodiments 1-12, wherein the polynucleotide comprises 1 to 50 PBS sequences.
14. The DNA repair complex of one of embodiments 1-13, wherein one or more PBS sequences comprise the nucleotide sequence of SEQ ID NO: 83.
15. The DNA repair complex of one of embodiments 1-14, wherein the PUF domain comprises a PUFa domain, a PUFb domain, a PUFc domain, or a PUFw domain.
16. The DNA repair complex of embodiment 15, wherein the PUFa domain has the amino acid sequence of SEQ ID NO: 27.
17. The DNA repair complex of any one of embodiments 1-16, wherein the RNA-guided DNA endonuclease is encoded by a first polynucleotide.
18. The DNA repair complex of any one of embodiments 1-17, wherein the polynucleotide is encoded by a second polynucleotide.

19. The DNA repair complex of any one of embodiments 1-18, wherein the DNA repair protein conjugate is encoded by a third polynucleotide.
20. The DNA repair complex of any one of embodiments 2-19, wherein the donor sequence is encoded by a fourth polynucleotide.
21. The DNA repair complex of embodiment 17, wherein the first polynucleotide is contained within a first vector.
22. The DNA repair complex of embodiment 18, wherein the second polynucleotide is contained within a second vector.
23. The DNA repair complex of embodiment 19, wherein the third polynucleotide is contained within a third vector.
24. The DNA repair complex of embodiment 20, wherein the fourth polynucleotide is contained within a fourth vector.
25. The DNA repair complex of one of embodiments 21-24, wherein either the first, second, third or fourth vector is the same.
26. The DNA repair complex of one of embodiments 21-25, wherein the first, second, third or fourth vector is a transfection vector.
27. The DNA repair complex of one of embodiments 21-25, wherein the first, second, third or fourth vector is a viral vector.
28. A cell comprising a DNA repair complex of one of embodiments 1-27.
29. The cell of embodiment 28, wherein the cell is a mammalian cell.
30. A method of increasing homology directed repair (HDR) in a mammalian cell, the method comprising:
 - (a) providing a mammalian cell containing a target nucleic acid requiring homology directed repair;
 - (b) delivering to the mammalian cell a first polynucleotide encoding an RNA-guided DNA endonuclease;
 - (c) delivering to the mammalian cell a second polynucleotide comprising:
 - (i) a DNA-targeting sequence that is complementary to a target polynucleotide sequence;
 - (ii) a binding sequence for the RNA-guided DNA endonuclease enzyme; and
 - (iii) one or more PUF binding site (PBS) sequences,wherein the RNA-guided DNA endonuclease enzyme is capable of binding to the second polynucleotide via the binding sequence;

(d) delivering to the mammalian cell a third polynucleotide comprising a DNA repair protein conjugate comprising:

(i) a PUF domain; and

(ii) a DNA repair domain, the DNA repair domain operably linked to the PUF domain; and

(e) delivering to the mammalian cell a fourth polynucleotide encoding a donor sequence complementary to the DNA-targeting sequence;

wherein the delivered DNA repair protein conjugate increases homology directed repair at the target nucleic acid sequence in the cell relative to a control.

31. The method of embodiment 30, wherein the DNA repair protein conjugate is bound to the second polynucleotide via binding of the PUF domain to the one or more PBS sequences.

32. The method of embodiment 30 or 31, wherein the first polynucleotide is contained within a first vector.

33. The method of one of embodiments 30-32, wherein the second polynucleotide is contained within a second vector.

34. The method of one of embodiments 30-33, wherein the third polynucleotide is contained within a third vector.

35. The method of one of embodiments 30-34, wherein the fourth polynucleotide is contained within a fourth vector.

36. The method of one of embodiments 30-35, wherein either the first, second, third or fourth vector are the same.

37. The method of one of embodiments 30-36, wherein the delivering is performed by transfection.

38. The method of one of embodiments 30-37, wherein the delivered DNA repair protein conjugate is capable of decreasing non-homologous end joining (NHEJ) at the target nucleic acid sequence in the cell relative to a standard control.

39. A method of decreasing non-homologous end joining (NHEJ) in a mammalian cell, the method comprising:

(a) providing a mammalian cell containing a target nucleic acid requiring NHEJ;

(b) delivering to the mammalian cell a first polynucleotide encoding an RNA-guided DNA endonuclease;

(c) delivering to the mammalian cell a second polynucleotide comprising:

(i) a DNA-targeting sequence that is complementary to a target polynucleotide sequence;

(ii) a binding sequence for the RNA-guided DNA endonuclease; and

(iii) one or more PUF binding site (PBS) sequences, wherein the RNA-guided DNA endonuclease is capable of binding to the second polynucleotide via the binding sequence;

(d) delivering to the mammalian cell a third polynucleotide encoding a DNA repair protein conjugate comprising:

(i) a PUF domain; and

(ii) a DNA repair domain, the DNA repair domain operably linked to the PUF domain; and

(e) delivering to the mammalian cell a fourth polynucleotide comprising a donor sequence complementary to the DNA-targeting sequence;

wherein the delivered DNA repair protein conjugate decreases non-homologous end joining (NHEJ) at the target nucleic acid sequence in the cell relative to a control.

40. The method of embodiment 39, wherein the DNA repair protein conjugate is bound to the second polynucleotide via binding of the PUF domain to the one or more PBS sequences.

41. The method of embodiment 39 or 40, wherein the first polynucleotide is contained within a first vector.

42. The method of one of embodiments 39-41, wherein the second polynucleotide is contained within a second vector.

43. The method of one of embodiments 39-42, wherein the third polynucleotide is contained within a third vector.

44. The method of one of embodiments 39-43, wherein the fourth polynucleotide is contained within a fourth vector.

45. The method of one of embodiments 39-44, wherein either the first, second, third or fourth vector are the same.

46. The method of one of embodiments 39-45, wherein the delivering is performed by transfection.

47. The method of one of embodiments 39-46, wherein the delivered DNA repair protein conjugate is capable of increasing HDR at the target nucleic acid sequence in the cell relative to a standard control.

48. A kit comprising:

(i) a ribonucleoprotein complex of one of embodiments 1-16 or a nucleic acid encoding the same; and

(ii) a DNA repair protein conjugate of one of embodiments 1-16 or a nucleic acid encoding the same.

49. The kit of embodiment 48, further comprising a transfection agent.
50. The kit of embodiment 48 or 49, further comprising a sample collection device for collecting a sample from a patient.

EXAMPLES

Example 1. Local recruitment of DNA repair proteins to enhance precise genome editing. DNA breaks are repaired through competing pathways containing overlapping and yet distinct protein components. Genetic studies in model organisms and human cells show that different templates require different pathways and protein factors. Drugs that perturb repair pathways have been applied to enhance genome editing with limited success, but these agents may also induce unwanted genomic instability. We took a more direct approach, which was to recruit DNA repair proteins locally to target sites. We developed a hybrid system based on CRISPR/Cas9 and the programmable Pumilio RNA-binding protein, termed “Casilio,” to recruit effector proteins at genomic targets. Here, we used direct Cas9 fusion or employed our Casilio technology to recruit DNA repair proteins or their complexes at the nuclease or nickase target site to favor homology-directed repair (HDR) of the double strand break (DSB) or single stranded nick (FIG. 2 and FIG. 4).

Example 2. Reporter cell lines for genome editing outcomes. For high-throughput quantification of genome editing outcomes and cell isolation by flow cytometry or fluorescence-activated cell sorting (FACS), we constructed a HDR/NHEJ reporter HEK293T cell line (FIG. 1A) with a constitutively expressed BFP gene inserted into the AAVS1 locus. Editing experiments were done by co-delivering Cas9 and a sgRNA targeting BFP with a repair template containing a H66Y mutation (single-stranded oligonucleotides, ssODN; or plasmid donor) that changes BFP to GFP. This reporter system conveniently reports on the fraction of cells that have undergone HDR (GFP-positive population), NHEJ (GFP-, BFP- double negative) and no modification (BFP-positive) (FIG. 1B).

Example 3. Recruitment of BRCA1 to site of Cas9-mediated double-stranded break enhances HDR. To test whether we can locally recruit BRCA1 to enhance HDR by direct tethering to Cas9 or through recruitment by Casilio, we fused BRCA1 to Cas9 or PUFa and tested the complexes’ abilities to enhance HDR in the context of the BFP->GFP reporter system. Tethering of BRCA by direct fusion to Cas9 N-terminus or C-terminus resulted in small decreases in HDR and decreased in NHEJ to greater extents. The total decrease in editing efficiency (HDR%+NHEJ%) may be due to steric disadvantages imposed on Cas9 by the large

BRCA1 protein (FIG. 3). Interestingly, tethering of BRCA1 through sgRNA scaffold mediated by Casilio stimulated HDR by 1.64- and 1.58-folds compared to Cas9-only control without compromising total editing efficiency (FIG. 5). These results indicate that tethering of a DNA repair protein can modulate genome editing pathway choice, and that the use of Casilio may be more effective than direct Cas9 fusion as a strategy for recruiting DNA repair proteins to enhance HDR.

Example 4. Recruitment of RAD54L to site of Cas9-mediated double-stranded break enhances HDR. We also recruited RAD54L to Cas9-DSB by fusing RAD54L to PUFa and recruiting it through Casilio-sgRNA scaffold. Such recruitment stimulated HDR by 1.31-fold with similar total editing efficiency (FIG. 6), demonstrating that recruitment of RAD54L by Casilio enhances HDR at Cas9 cut site without compromising total editing efficiency.

Example 5. Recruitment of CtIP(T847E)-PALB2(KR)-BRCA1 to site of Cas9-mediated double-stranded break enhances HDR. The complex formation capabilities of Casilio do not only allow the recruitment of individual proteins, but also the assembly of multiprotein complexes—either multiple molecules of a particular protein or combinations of different proteins—at the target site. We thus tested whether we can recruit multiple DNA repair proteins to the site of Cas9-mediated double-stranded break by the Casilio approach (FIG. 7). We transfected HEK293T/BFP reporter cell with Cas9, BFP->GFP ssODN, sgBFP-5xPBSa and a cocktail of CtIP(T847E)-PUFa, PALB2(KR)-PUFa and BRCA1-PUFa fusion and found that such complex stimulated HDR by 1.63-fold (FIG. 8). Similar C-terminal fusions of the three proteins to PUFa stimulated HDR by 1.28-fold. This example demonstrates the utility of Casilio to recruit multiple DNA repair proteins to the DSB target site to enhance HDR.

Example 6. Recruitment of RAD51A enhances HDR at site of DNA nick mediated by Cas9Nickase (Cas9 D10A nickase) enhances HDR. We sought to explore whether recruitment of DNA repair proteins in the context of single-stranded nicks mediated by Cas9Nickase (Cas9n) would enhance HDR while keeping minimal NHEJ. We fused RAD51A to PUFa N-terminally (RAD51A-PUFa). We then transfected cells with Cas9 Nickase, BFP->GFP ssODN, sgBFP-5xPBSa and RAD51A-PUFa into HEK293T/BFP reporter cell and observed 3.55-fold stimulation of HDR (FIG. 9).

Example 7. Recruitment of CtIP(T847E)-PALB2(KR)-BRCA1 at site of DNA nick mediated by Cas9Nickase (Cas9 D10A nickase) enhances HDR. We tested whether the recruitment of CtIP(T847E)-PALB2(KR)-BRCA1 to Cas9 Nickase (Cas9 D10A nickase) would enhance HDR. We transfected HEK293T/BFP reporter cell with Cas9Nickase (Cas9n), BFP-

>GFP ssODN, sgBFP-5xPBSa and a cocktail of CtIP(T847E)-PUFa, PALB2(KR)-PUFa and BRCA1-PUFa fusions and found that such complex stimulated HDR by 2.87-fold (FIG. 10). Similar C-terminal fusions of the three proteins to PUFa also stimulated HDR by 2.84-fold. This example demonstrates the utility of Casilio to recruit multiple DNA repair proteins to the nick site mediated by Cas9Nickase to enhance HDR.

Example 8. Recruitment of XRCC3, RECQ5 or FEN1 to site of Cas9-mediated double-stranded break enhances HDR. We also recruited XRCC3, RECQ5, or FEN1 to Cas9-DSB by fusing each of them N- or C-terminally to PUFa to allow local recruitment via Casilio-sgRNA scaffold. Recruitment of XRCC3 stimulated HDR by twofold (XRCC3-PUFa) or 1.84-fold (PUFa-XRCC3) (FIG. 11). Recruitment of RECQ5 stimulated HDR by 1.85-fold (RECQ5-PUFa) or 1.58-fold (PUFa-RECQ5) (FIG. 12). Recruitment of FEN1 stimulated HDR by 1.97-fold (FEN1-PUFa) or 1.86-fold (PUFa-FEN1) (FIG. 13).

Example 9. Recruitment of Fanconi Anemia (FA) pathway proteins to site of Cas9-mediated double-stranded break enhances HDR. We also recruited proteins of the Fanconi Anemia (FA) pathway (FANCB, FANCF, FANCG, FANCM) to Cas9-DSB by fusing each of them to PUFa to allow local recruitment via Casilio-sgRNA scaffold. Recruitment of FANCB stimulated HDR by 1.76-fold (PUFa-FANCB) (FIG. 14). Recruitment of FANCF stimulated HDR by 2.35-fold (FANCF-PUFa) or 2.36-fold (PUFa-FANCF); Recruitment of FANCG stimulated HDR by 2.07-fold (FANCG-PUFa) or 2.15-fold (PUFa-FANCG); Recruitment of FANCM stimulated HDR by 2.12-fold (FANCM-PUFa) or 1.79-fold (PUFa-FANCM) (FIG. 14).

Example 10. More examples of factors that enhance HDR when recruited to site of Cas9Nickase (Cas9 D10A nickase)-mediated DNA nick. We also recruited more factors to Cas9Nickase (Cas9 D10A nickase)-mediated nick via Casilio-sgRNA scaffold (FIG. 15). Recruitment of PALB2 stimulated HDR by 33%. Recruitment of PALB2(KR) stimulated HDR by 50%; Recruitment of MRE11A stimulated HDR by 58%; Recruitment of PALB2(delETGE) stimulated HDR by 75%; Recruitment of FANCG stimulated HDR by 2-fold; Recruitment of CtIP(T847E) stimulated HDR by 2.17-fold; Recruitment of UBE2T stimulated HDR by 2.25-fold; Recruitment of FEN1 stimulated HDR by 2.33-fold; Recruitment of CtIP stimulated HDR by 2.42-fold; Recruitment of RAD54L stimulated HDR by 2.5-fold; Recruitment of USP1 stimulated HDR by 2.5-fold; Recruitment of XRCC3 stimulated HDR by 2.58-fold; Recruitment of RPA2 stimulated HDR by 2.58-fold; Recruitment of BRC3 stimulated HDR by 3.08-fold; Recruitment of FANCM stimulated HDR by 3.42-fold; Recruitment of RPA3 stimulated HDR

by 3.5-fold; Recruitment of BRC4 stimulated HDR by 3.58-fold; Recruitment of RPA1 stimulated HDR by 4.08-fold.

REFERENCES

1. Khan, F.A. et al. CRISPR/Cas9 therapeutics: a cure for cancer and other genetic diseases. *Oncotarget* **7**, 52541-52552 (2016).
2. Xiong, X., Chen, M., Lim, W.A., Zhao, D. & Qi, L.S. CRISPR/Cas9 for Human Genome Engineering and Disease Research. *Annual review of genomics and human genetics* **17**, 131-154 (2016).
3. Wiles, M.V., Qin, W., Cheng, A.W. & Wang, H. CRISPR-Cas9-mediated genome editing and guide RNA design. *Mammalian genome : official journal of the International Mammalian Genome Society* **26**, 501-510 (2015).
4. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* **5**, a012740 (2013).
5. Orr-Weaver, T.L. & Szostak, J.W. Yeast recombination: the association between double-strand gap repair and crossing-over. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 4417-4421 (1983).
6. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E924-932 (2014).
7. Davis, L. & Maizels, N. DNA nicks promote efficient and safe targeted gene correction. *PLoS One* **6**, e23981 (2011).
8. Kan, Y., Ruis, B., Takasugi, T. & Hendrickson, E.A. Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res* **27**, 1099-1111 (2017).
9. Bothmer, A. et al. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun* **8**, 13905 (2017).
10. Paix, A. et al. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E10745-e10754 (2017).
11. Richardson, C.D. et al. CRISPR-Cas9 Genome Editing In Human Cells Works Via The Fanconi Anemia Pathway. *bioRxiv* (2017).

12. Chu, V.T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-548 (2015).
13. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**, 538-542 (2015).
14. Li, G. et al. Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. *Sci Rep* **7**, 8943 (2017).
15. Yu, C. et al. Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* **16**, 142-147 (2015).
16. Takayama, K. et al. Highly efficient biallelic genome editing of human ES/iPS cells using a CRISPR/Cas9 or TALEN system. *Nucleic Acids Res* **45**, 5198-5207 (2017).
17. Charpentier, M. et al. CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat Commun* **9**, 1133 (2018).
18. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* **31**, 230-232 (2013).
19. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012-1019 (2014).
20. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F. & Chin, L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell reports* **14**, 1555-1566 (2016).
21. Lin, S., Staahl, B.T., Alla, R.K. & Doudna, J.A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).
22. Guo, Q. et al. 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. *Sci Rep* **8**, 2080 (2018).
23. Ma, M. et al. Efficient generation of mice carrying homozygous double-floxed alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Res* **27**, 578-581 (2017).
24. Jasin, M. & Haber, J.E. The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA repair* **44**, 6-16 (2016).
25. Cheng, A.W. et al. Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. *Cell research* (2016).

26. Streeter, I. et al. The human-induced pluripotent stem cell initiative-data resources for cellular genetics. *Nucleic Acids Res* **45**, D691-d697 (2017).
27. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. & Corn, J.E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-344 (2016).
28. Certo, M.T. et al. Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods* **8**, 671-676 (2011).
29. Kuhar, R. et al. Novel fluorescent genome editing reporters for monitoring DNA repair pathway utilization at endonuclease-induced breaks. *Nucleic Acids Res* **42**, e4 (2014).

What is claimed is:

1. A method comprising delivering to a cell comprising a target nucleic acid:
 - (a) a ribonucleic acid (RNA)-guided nuclease or a nucleic acid encoding a RNA-guided nuclease;
 - (b) a guide RNA (gRNA) or a nucleic acid encoding a gRNA that comprises (i) a DNA-targeting sequence that can bind specifically to a target nucleic acid sequence, (ii) a RNA-guided nuclease-binding sequence, and (iii) a Pumilio-FBF (PUF) domain binding sequence (PBS); and
 - (c) a DNA repair conjugate or a nucleic acid encoding a DNA repair conjugate that comprises a DNA repair domain linked to a PUF domain that binds to the PBS.
2. The method of claim 1, wherein the target nucleic acid sequence is in the genome of the cell.
3. The method of any one of the preceding claims, wherein the target nucleic acid sequence is within a gene and/or within a transcriptional regulatory sequence.
4. The method of any one of the preceding claims, wherein the method comprises delivering to the cell a ribonucleoprotein complex comprising the RNA-guided nuclease bound to the gRNA.
5. The method of any one of the preceding claims further comprising delivering to the cell a donor nucleic acid comprising a sequence of interest.
6. The method of claim 5, wherein the donor nucleic acid is single-stranded or double-stranded.
7. The method of any one of the preceding claims further comprising maintaining the cell under conditions that result in cleavage of the target nucleic acid sequence.
8. The method of claim 7 further comprising maintaining the cell under conditions that result in the production of a cellular nucleic acid comprising the sequence of interest.

9. The method of any one of the preceding claims, wherein the RNA-guided DNA nuclease is a Cas9 nuclease or a Cas9 nickase.
10. The method of any one of the preceding claims, wherein the gRNA comprises at least two PUF domain binding sequences.
11. The method of claim 10, wherein the gRNA comprises 2-50 PUF domain binding sequences.
12. The method of claim 10 or 11, wherein at least two of the PUF binding domain sequences are tandem repeat sequences.
13. The method of any one of the preceding claims, wherein the PBS has a length of at least 8 nucleotides.
14. The method of any one of the preceding claims, wherein the PBS comprises the nucleotide sequence of SEQ ID NO: 83.
15. The method of any one of the preceding claims, wherein the PUF domain comprises a PUFa domain, a PUFb domain, a PUFc domain, or a PUFw domain.
16. The method of claim 15, wherein the PUF domain comprises a PUFa domain that comprises the amino acid sequence of SEQ ID NO: 27.
17. The method of any one of the preceding claims, wherein the PUF domain comprises a C-terminus and a N-terminus, and wherein the DNA repair domain is linked to the C-terminus or the N-terminus of the PUF domain.
18. The method of any one of the preceding claims, wherein the DNA repair domain comprises an enzymatic activity selected from the group consisting of ligase activity, polymerase activity, topoisomerase activity, helicase activity, and nuclease activity.
19. The method of any one of the preceding claims, wherein the DNA repair domain comprises a protein selected from the group consisting of: Replication Protein A1 (RPA1); Replication Protein A2 (RPA2); Fanconi Anemia Complementation Group M (FANCM); RAD51 Recombinase (RAD51); RAD52 Homolog, DNA Repair Protein (RAD52); RAD51

Paralog C (RAD51C); RAD18 E3 Ubiquitin Protein Ligase (RAD18); RB Binding Protein 8, Nuclease (RBBP8/CTIP); Tumor Protein P53 Binding Protein 1 (TP53BP1); BRCA1 DNA Repair Associated (BRCA1); RAD54-like (RAD54L); Partner and Localizer of BRCA2 (PALB2); X-Ray Repair Cross Complementing 3 (XRCC3); MRE11 Homolog, Double Strand Break Repair Nuclease (MRE11A); Flap Structure-Specific Nuclease 1 (FEN1); RecQ Like Helicase 5 (RECQ5); FA Complementation Group B (FANCB); Ubiquitin Specific Peptidase 1 (USP1); FA Complementation Group F (FANCF); and FA Complementation Group G (FANCG).

20. The method of any one of the preceding claims, wherein the gRNA, the RNA-guided nuclease, the DNA repair conjugate, and/or the donor nucleic acid are encoded on independent vectors or on the same vector.

21. The method of claim 20, wherein the vector is a plasmid or a viral vector.

22. A DNA repair conjugate comprising a DNA repair domain linked to a Pumilio-FBF (PUF) domain.

23. The DNA repair conjugate of claim 22, wherein the DNA repair domain comprises an enzymatic activity selected from the group consisting of ligase activity, polymerase activity, topoisomerase activity, helicase activity, and nuclease activity.

24. The DNA repair conjugate of claim 22 or 23, wherein the DNA repair domain comprises a protein selected from the group consisting of: RPA1; RPA2; FANCM; RAD51; RAD 52; RAD51C; RAD18; RBBP8/CTIP; TP53BP1; BRCA1; RAD54L; PALB2; XRCC3; MRE11A; FEN1; RECQ5; FANCB; USP1; FANCF; and FANCG.

25. The DNA repair conjugate of any one of the preceding claims, wherein the PUF domain comprises a PUFa domain, a PUFb domain, a PUFc domain, or a PUFw domain.

26. The DNA repair conjugate of claim 25, wherein the PUF domain comprises a PUFa domain that comprises the amino acid sequence of SEQ ID NO: 27.

27. A nucleic acid encoding the DNA repair conjugate of any one of the preceding claims.

28. An expression vector comprising a promoter operably linked to the nucleic acid of claim 27.

29. A kit comprising the DNA repair conjugate, the nucleic acid, or the expression vector of any one of the preceding claims.

30. The kit of claim 29 further comprising a RNA-guided nuclease or a nucleic acid encoding a RNA-guided nuclease.

31. The kit of claim 29 or 30 further comprising a gRNA or a nucleic acid encoding a gRNA that comprises (i) a DNA-targeting sequence that is complementary to a target nucleic acid sequence, (ii) a RNA-guided nuclease-binding sequence, and (iii) a PBS, wherein the PUF domain of the DNA repair conjugate can bind to the PBS.

32. The kit of any one of the preceding claims further comprising a donor nucleic acid that comprises a sequence of interest.

33. A cell comprising the DNA repair conjugate, the nucleic acid, or the expression vector of any one of the preceding claims.

34. The cell of claim 33 further comprising a RNA-guided nuclease or a nucleic acid encoding a RNA-guided nuclease.

35. The cell of claim 34, wherein the RNA-guided nuclease is a Cas9 nuclease or Cas9 nickase.

36. The cell of claim 34 or 35 further comprising a gRNA or a nucleic acid encoding a gRNA that comprises (i) a DNA-targeting sequence that is complementary to a target nucleic acid sequence, (ii) a RNA-guided nuclease-binding sequence, and (iii) a PBS to which the PUF domain of the DNA repair conjugate can bind.

37. A method comprising delivering to a cell a programmable nuclease-based gene editing system that comprises a programmable nuclease linked to a DNA repair domain, wherein the programmable nuclease cleaves the target nucleic acid sequence, and the DNA repair domain is selected from the group consisting of: RPA1; RPA2; FANCM; BRCA1; RAD54L; PALB2; XRCC3; FEN1; RECQ5; FANCB; USP1; FANCF; and FANCG.

38. The method of claim 37 further comprising delivering to the cell a donor nucleic acid comprising a sequence of interest.
39. The method of claim 37 or 38, wherein the programmable nuclease comprises a RNA-guided nuclease.
40. The method of claim 39, wherein the RNA-guided nuclease is a Cas9 nuclease or a Cas 9 nickase.
41. The method of claim 39 or 40 further comprising delivering to the cell a gRNA or a nucleic acid encoding a gRNA that specifically binds to the target nucleic acid sequence.
42. The method of claim 37 or 38, wherein the programmable nuclease comprises a zinc finger nuclease (ZFN).
43. The method of claim 37 or 38, wherein the programmable nuclease comprises a transcription activator-like effector nuclease (TALEN).
44. The method of any one of claims 37-43, wherein the programmable nuclease is indirectly linked to the DNA repair domain.
45. A DNA repair protein comprising a programmable nuclease linked to DNA repair domain selected from the group consisting of: RPA1; RPA2; FANCM; BRCA1; RAD54L; PALB2; XRCC3; FEN1; RECQ5; FANCB; USP1; FANCF; and FANCG.
46. The DNA repair protein of claim 45, wherein the programmable nuclease comprises a RNA-guided nuclease.
47. The DNA repair protein of claim 46, wherein the RNA-guided nuclease comprises Cas9 nuclease or Cas9 nickase.
48. The DNA repair protein of claim 45, wherein the programmable nuclease comprises a ZFN.

49. The DNA repair protein of claim 45, wherein the programmable nuclease comprises a TALEN.
50. The DNA repair protein of any one of claims 45-49, wherein the programmable nuclease is indirectly linked to the DNA repair domain.
51. A nucleic acid encoding the DNA repair protein of any one of claims 45-50.
52. An expression vector comprising a promoter operably linked to the nucleic acid of claim 51.
53. A kit comprising the DNA repair protein of any one of claims 45-50, the nucleic acid of claim 51, or the expression vector of claim 52.
54. The kit of claim 53 comprising the DNA repair protein of claim 45 or 46 and a gRNA or a nucleic acid encoding a gRNA that comprises (i) a DNA-targeting sequence that is complementary to a target nucleic acid sequence and (ii) a RNA-guided nuclease-binding sequence.
55. The kit of any claim 53 or 54 further comprising a donor nucleic acid that comprises a sequence of interest.

FIG. 1A

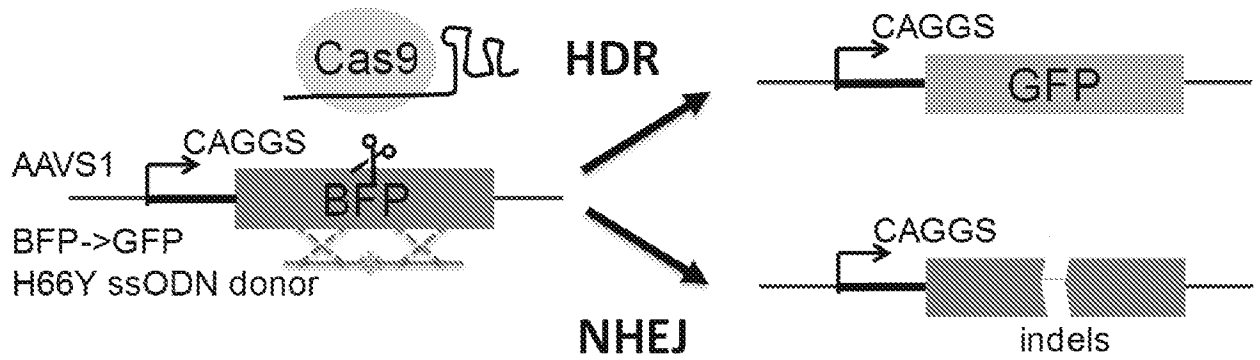


FIG. 1B

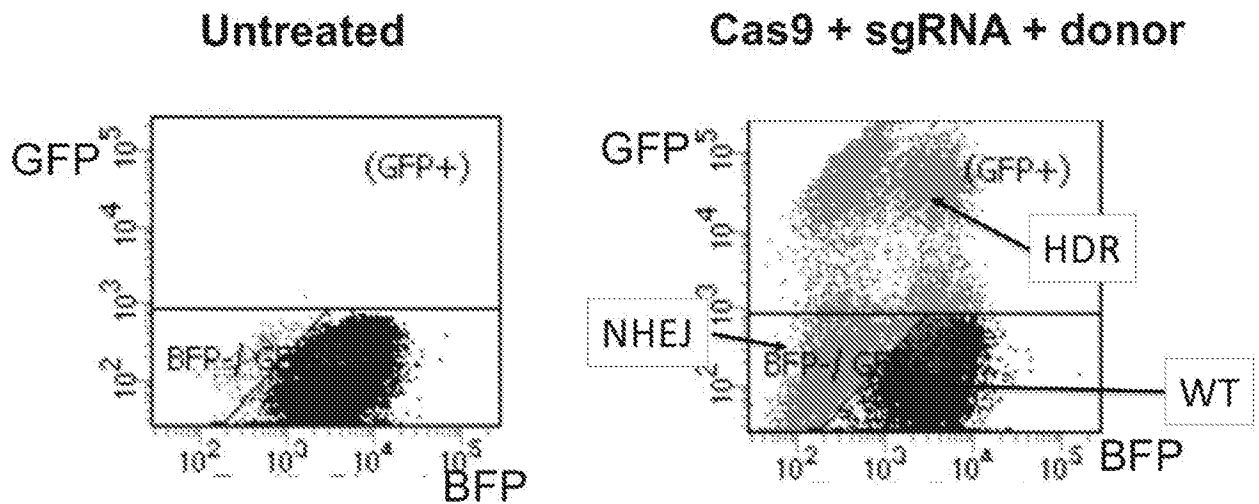
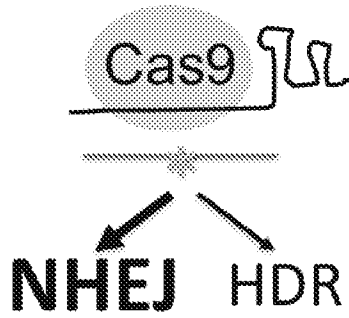


FIG. 2

(i) Unfused Cas9



(ii) Direct Cas9-DNA repair protein fusion

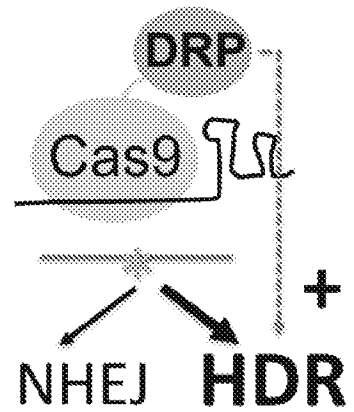


FIG. 3

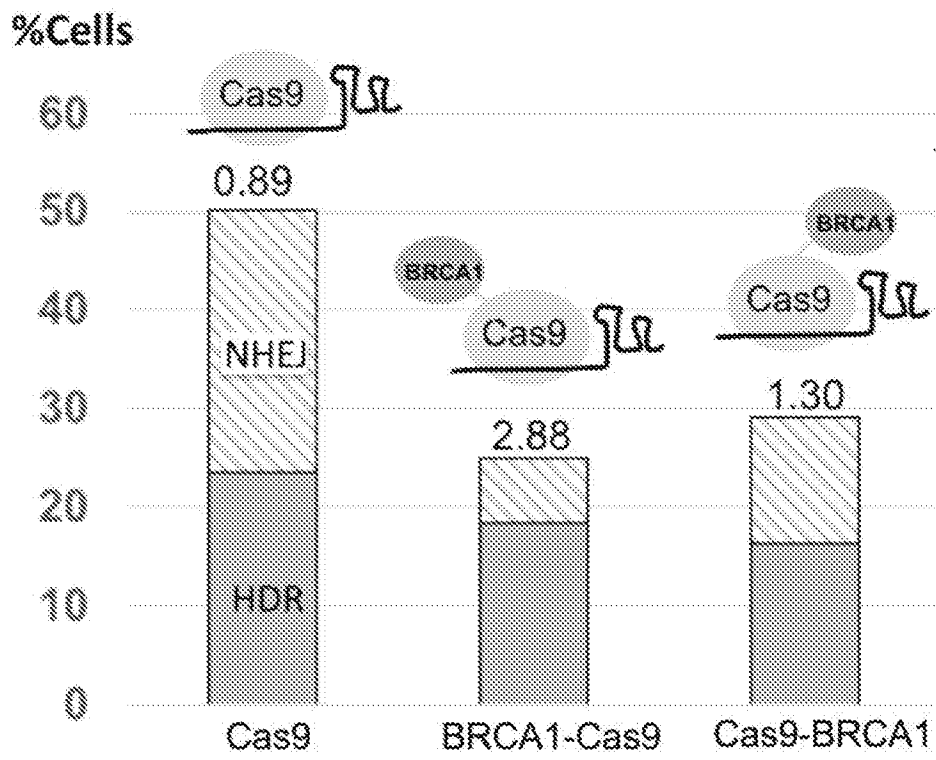


FIG. 4

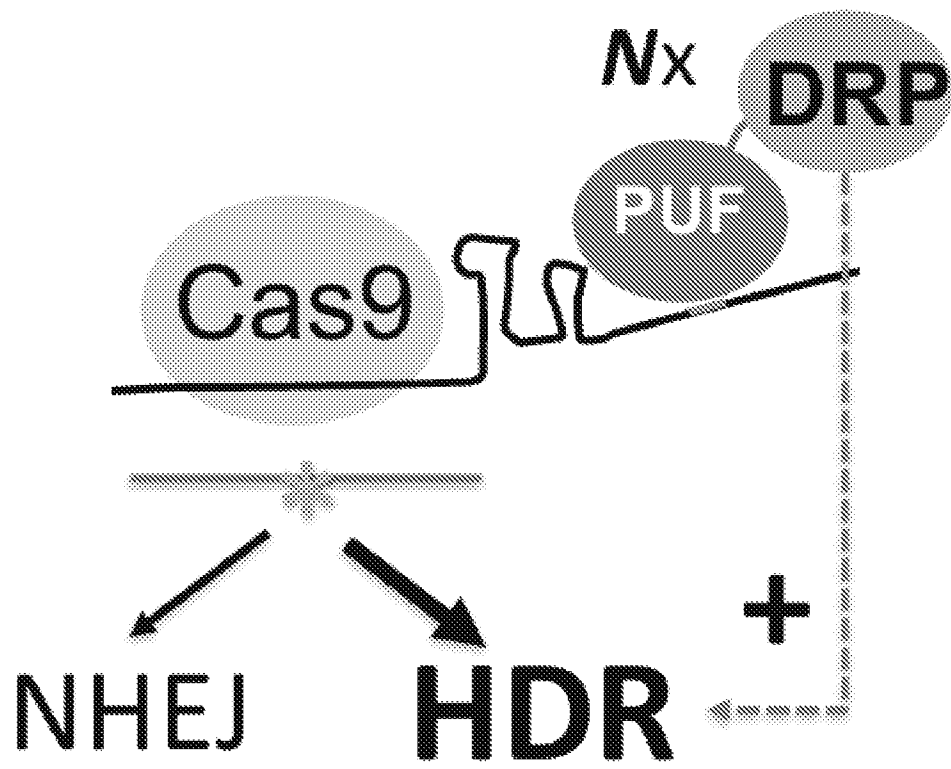


FIG. 5

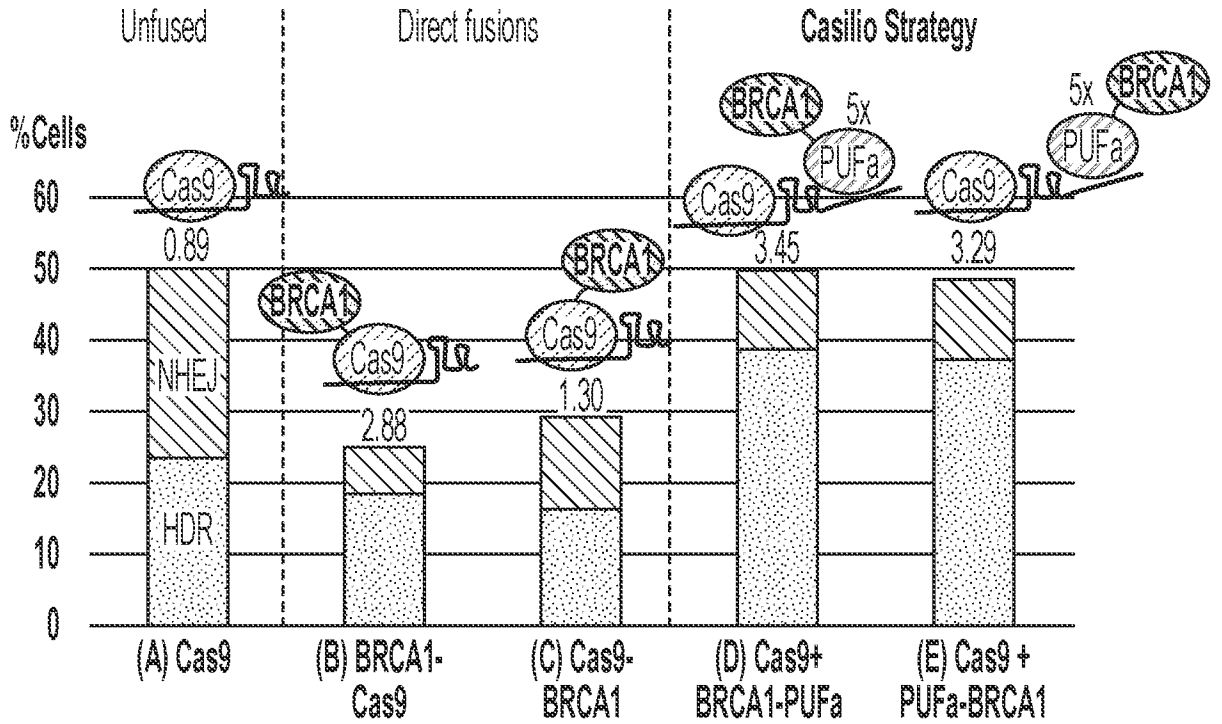


FIG. 6

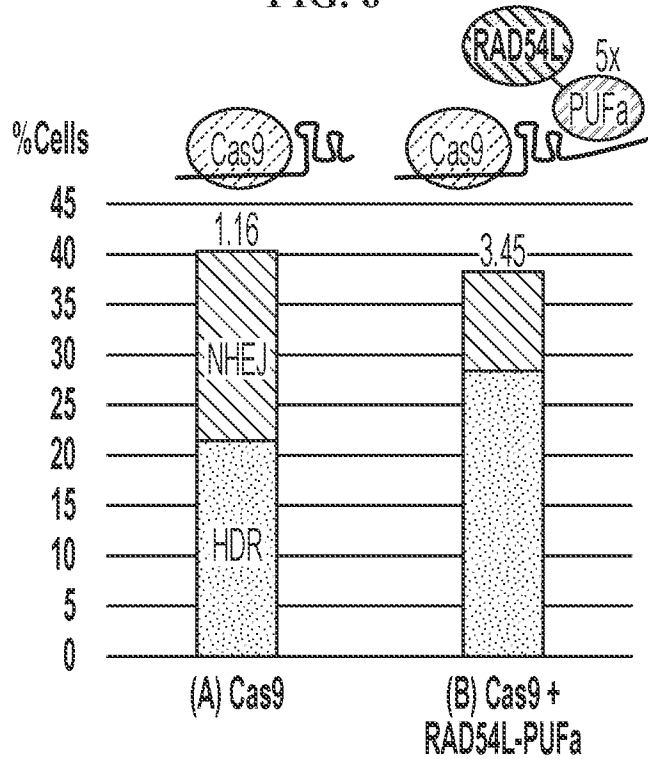


FIG. 7

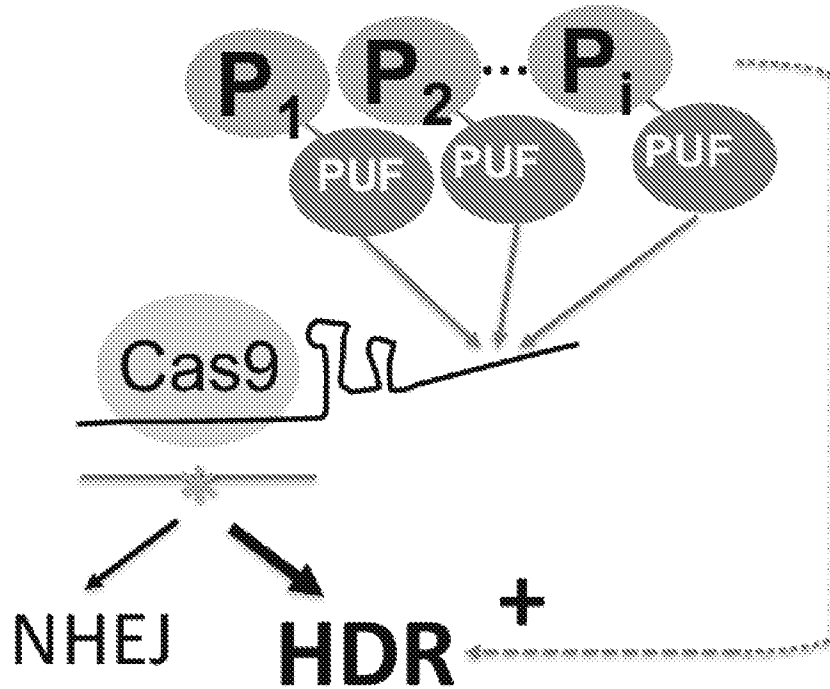


FIG. 8

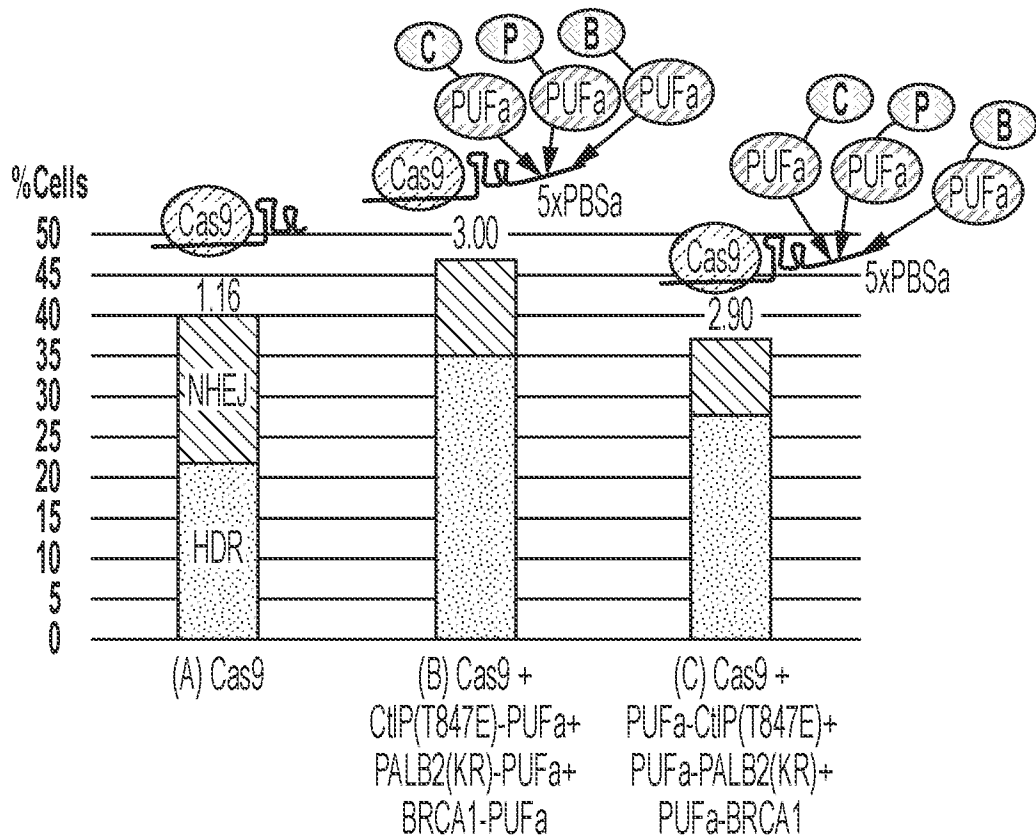


FIG. 9

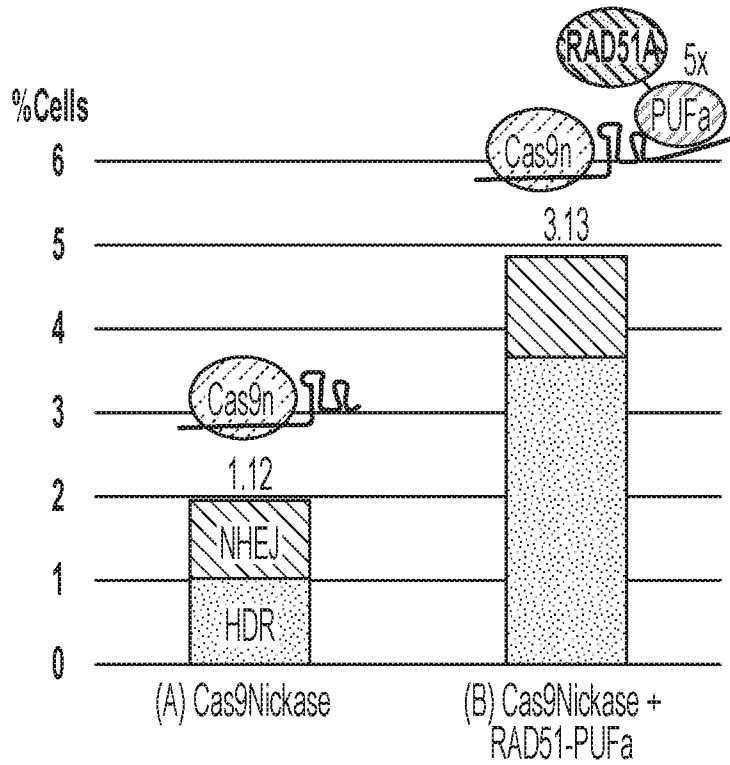


FIG. 10

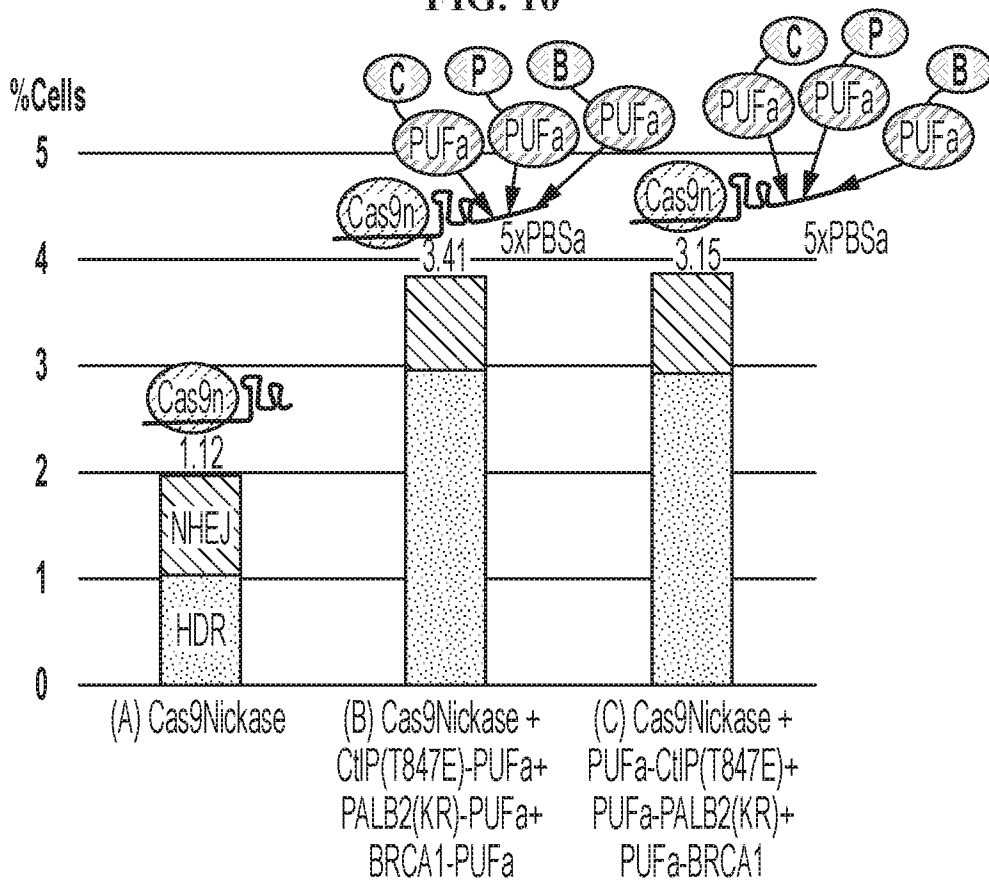


FIG. 11

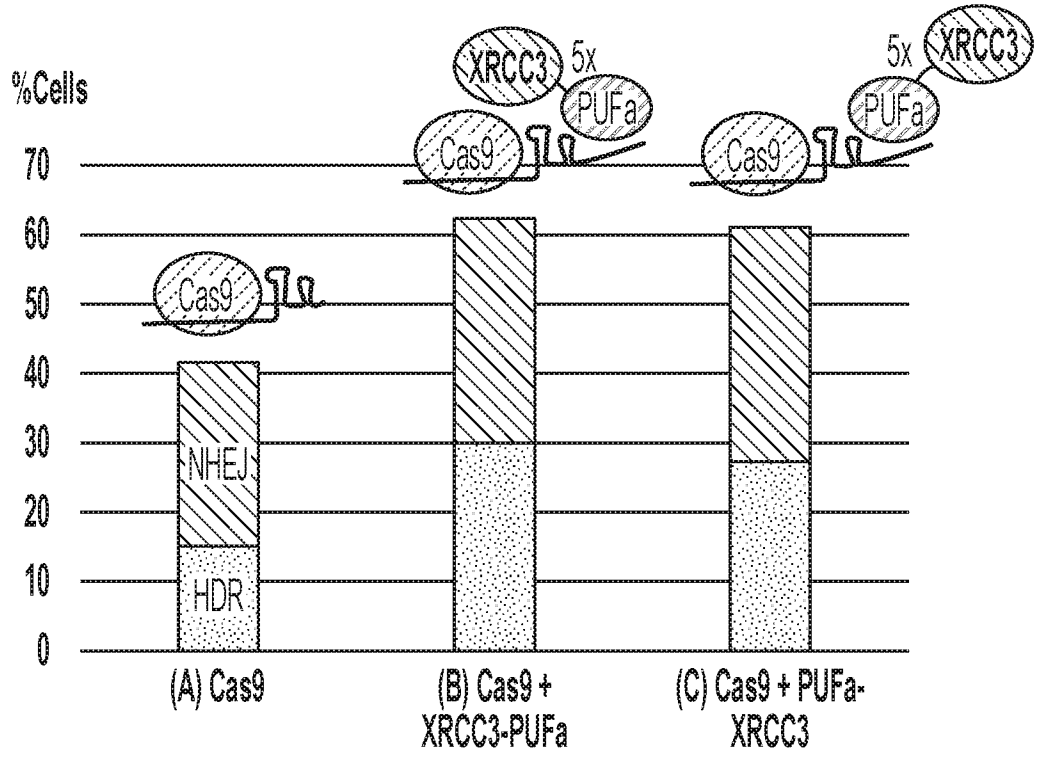


FIG. 12

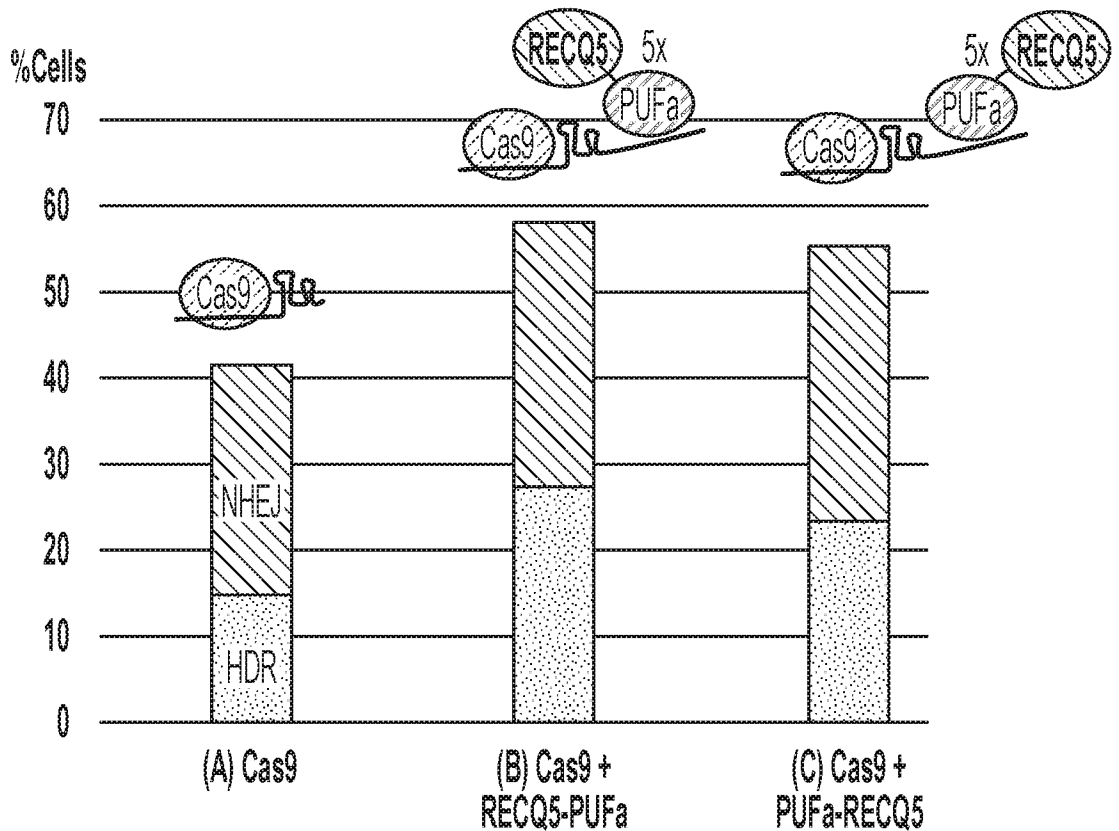


FIG. 13

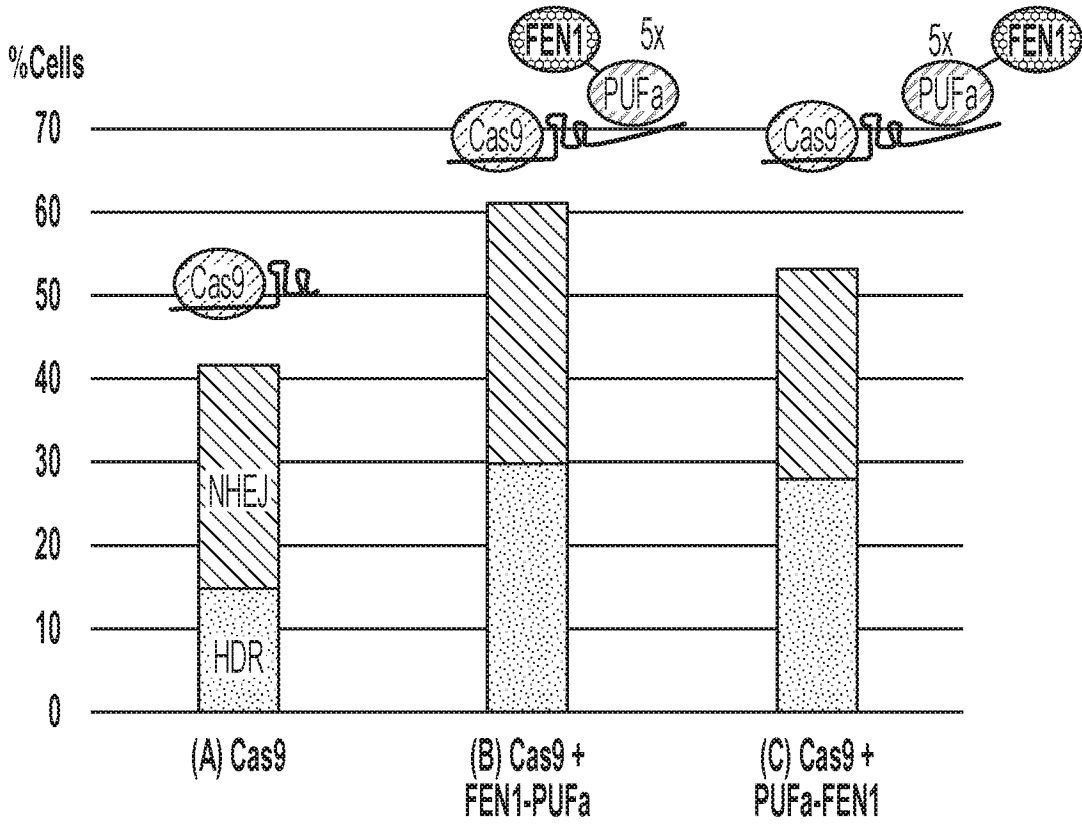
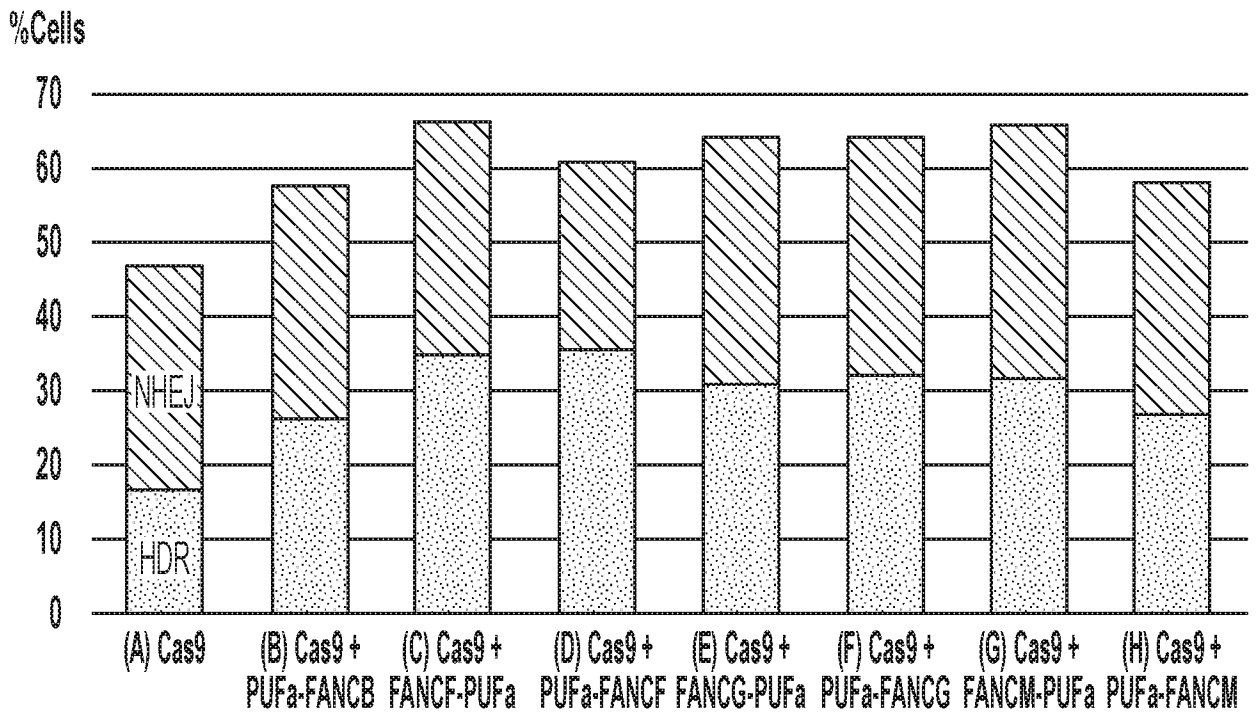
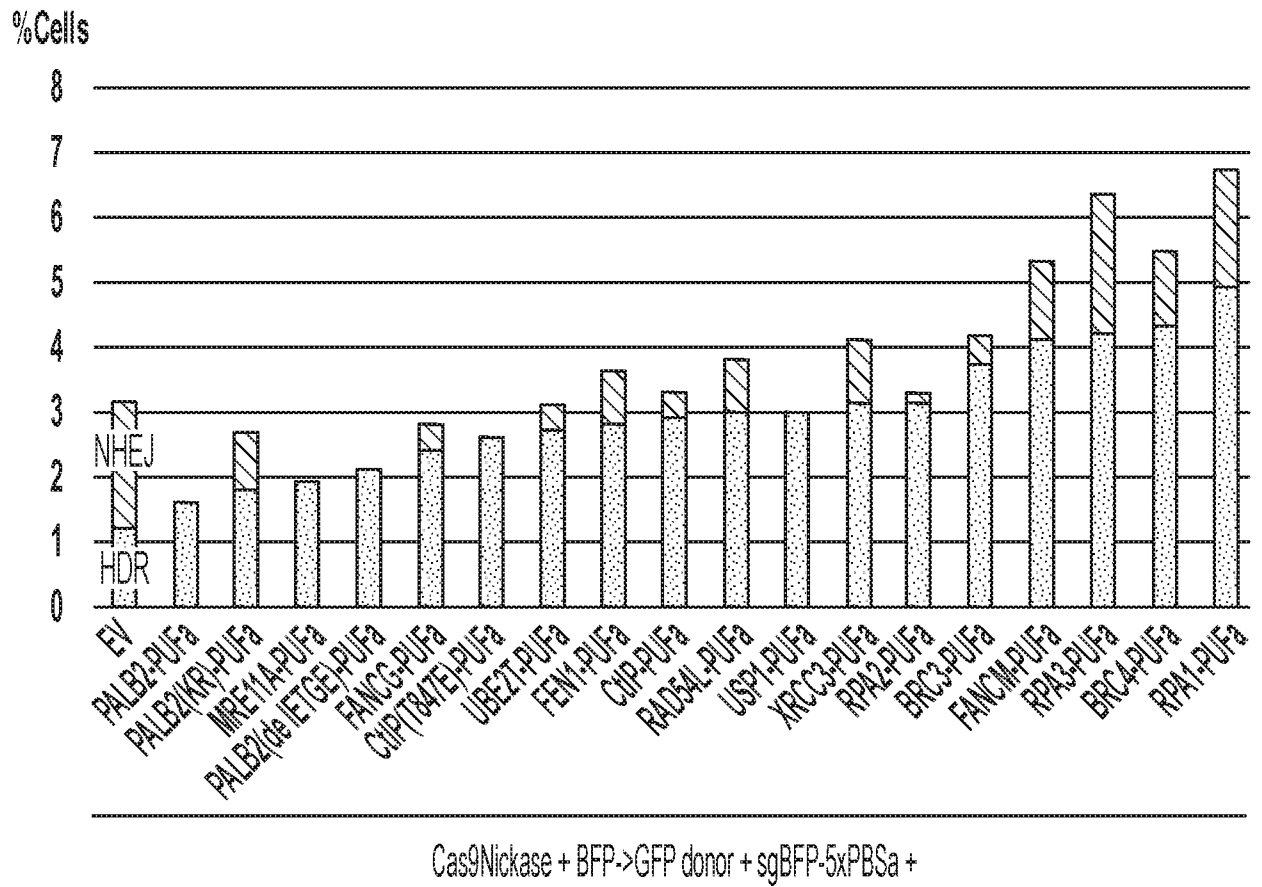


FIG. 14



10/10

FIG. 15



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/047021

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

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

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

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

- 3. Claims Nos.: 4-21, 25-36, 41, 44, 51-55
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

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

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/047021

A. CLASSIFICATION OF SUBJECT MATTER
 IPC(8) - C12N 9/22; C12N 9/96; C12N 15/90 (2019.01)
 CPC - C07K 2319/80; C12N 9/22; C12N 9/96; C12N 15/90; C12N 15/907; C12N 2310/20 (2019.08)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2018/0094257 A1 (JACKSON LABORATORY) 05 April 2018 (05.04.2018) entire document	1-3, 22 --- 23, 24
X --- Y	US 2018/0230494 A1 (GENERAL HOSPITAL CORPORATION) 16 August 2018 (16.08.2018) entire document	37-40, 42, 43, 45-50 --- 23, 24
A	CHENG et al, Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labeling. Cell Res. (February 2016, published online 15 January 2016) Vol. 26, No. 2, Pg. 254-7; entire document	1-3, 22-24, 37-40, 42, 43, 45-50
A	BOTHMER et al, Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. Nat Commun. (09 January 2017) Vol. 8, No. 13905, Pg. 1-12; entire document	1-3, 22-24, 37-40, 42, 43, 45-50
A	WO 2017/142923 A1 (EMENDOBIO) 24 August 2017 (24.08.2017) entire document	1-3, 22-24, 37-40, 42, 43, 45-50
A	PAULSEN et al, Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR-Cas9 genome editing. Nat Biomed Eng. (November 2017, published online 09 October 2017) Vol. 1, No. 11, Pg. 878-888; entire document	1-3, 22-24, 37-40, 42, 43, 45-50
P, A	WO 2018/162702 A1 (INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE (INSERM)) 13 September 2018 (13.09.2018) entire document	1-3, 22-24, 37-40, 42, 43, 45-50

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

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
 25 October 2019

Date of mailing of the international search report
19 NOV 2019

Name and mailing address of the ISA/US
 Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
 P.O. Box 1450, Alexandria, VA 22313-1450
 Facsimile No. 571-273-8300

Authorized officer
 Blaine R. Copenheaver
 PCT Helpdesk: 571-272-4300
 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/047021

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, A	REES et al, Development of hRad51-Cas9 nickase fusions that mediate HDR without double-stranded breaks. Nat Commun. (17 May 2019) Vol. 10, No. 1:2212, Pg. 1-12; entire document	1-3, 22-24, 37-40, 42, 43, 45-50