Methods and Compositions for Improvement of Base Editor and Homology Directed Repair Genome Editing Outcomes

Related Applications

[0001] The present application claims the benefit of priority of U.S. Provisional Application No. 63/573,216, filed April 2, 2024, the entire contents of which is incorporated herein by reference.

Sequence Listing

[0002] The present application contains a Sequence Listing that has been submitted in XML format via Patent Center and is hereby incorporated by reference in its entirety. The XML copy, created on February 5, 2025, is named P24-059-WO-PCT_SL, and is 43.4 kilobytes in size.

Background

[0003] DNA damage repair mechanisms present in eurkaryotic cells can be manipulated to enable the installation of precision edits within a target DNA sequence. When the cell detects damage to DNA, such as a single stranded nick or a double stranded break in the chromosomal sequence, there are cascading DNA damage responses from the cell that can be utilized for precision editing. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) base editing systems enable the installation of precision substitutions within a target DNA by catalytically converting one nucleobase into another. These systems are comprised of a CRISPR effector, such as Cas9 or Cas12, along with a fused enzymatic domain that performs the nucleobase conversion. For cytosine base editors, the fused enzymatic domain is a deaminase that catalyzes the conversion of cytosine to uracil, leading to C-to-T substitutions; for adenine base editors; the fused enzymatic domain is a deaminase that catalyzes the conversion of adenine to inosine, leading to A-to-G substitutions; for guanine base editors, the enzymatic domain is a glycosylase that excises the guanine nucleobase, leading to G-to-Y substitutions. The Cas effector directs the base editor molecule to the desired target DNA by base pairing

between the target and a complementary guide RNA bound to the Cas effector. The Cas effector is engineered to be partially or fully nuclease-inactive to prevent generation of double-strand breaks (DSBs) in the target DNA, thereby minimizing the generation of unwanted insertions and deletions.

[0004] The enzymatic conversion of nucleobases by base editors resembles DNA damage. Background spontaneous deamination of cytosine and adenine occurs at low but measurable rates and result in the installation of substitutions. Abasic sites result in the installation of substitutions and are prone to breakage of the phosphodiester backbone, leading to DNA breaks and, in some cases, insertions and deletions (indels). Accumulation of these substitutions and indels would be catastrophic for genome stability. Therefore, organisms have evolved DNA repair pathways to reverse these potentially mutagenic events. These repair pathways also can result in the reversion of a successfully edited site to its pre-edited state. For example, following cytosine base editing, the base excision repair (BER) pathway excises the uracil nucleobase from DNA to generate an abasic site and uses the complementary strand to direct the insertion of cytosine at the edited position. In some cases, DNA repair activity in combination with base editing can cause indels. Some versions of base editors utilize a nickase Cas effector. In the absence of BER, nicking of the unedited strand by Cas results in an increased rate of edit installation (Komor, *et al.*, Nature, 2016). However, breakage of fragile abasic sites that result from BER may lead to nicks on both strands of the target DNA; this DSB is most often repaired by nonhomologous end joining (NHEJ) and frequently results in indels. Cytosine base editors typically mitigate both of these undesired outcomes by the inclusion of uracil glycosylase inhibitor (UGI), which inhibits the excision of uracil from DNA to promote the installation of the desired C-to-T edit.

[0005] Others have attempted to improve base editing with other strategies but with limited success. Bae, *et al.*, have reported the genetic disruption of NHEJ by knocking out KU70, a protein utilized in NHEJ (Bae, *et al.*, Multiplex Gene Disruption by targeted Base Editing

of *Yarrowia Lipoytica* Genome Using Cytidine Deaminase Combined with the CRISPR/Cas9 System, Biotechnol. J., 2020 Jan:15(1)). However, this method is both difficult and time consuming.

[0006] Both Liu, *et al.* (Liu, *et al.*, HDAC Inhibitors Improve CRISPR-Cas9 Mediated Prime Editing and Base Editing, Mol. Ther. Nucleic Acids, 2022, Oct 06;30:173) and Zhao, *et al.*, (Zhao, *et al.*, Small Molecule Compounds Boost Genome-Editing Efficiency of Cytosine Base Editor, Nucleic Acids Res., 2021, Sept 07;49(15):8974-8986) detail the use of histone deacetylase (HDAC) inhibitors to improve base editor outcomes with some success.

[0007] Church, *et al.*, (US Patent Publication No. 2022/177877) discloses methods and compositions for multiplexed base editing. The disclosure requires inactive Cas9 and multiplex editing of repetitive elements of at least five targets simultaneously. While the disclosure describes inhibitors of base excision repair and inhibitors of NHEJ, their use is restricted to the specific, limited methods of the disclosure, which is indicative to one of skill in the art of a lack of success with other methods. Further, they did not disclose a single-strand break repair (SSBR) inhibitor for use in base editing.

[0008] Further improvements in the rate of C-to-T conversion, along with reductions in the rate of indel formation, would be beneficial to the field of genome editing. For therapeutic applications, in particular, a simultaneous increase in the rate of successful edit and decrease in the rate of undesired byproduct edits would increase the likelihood of success.

[0009] Furthermore, CRISPR editing enables precise editing when delivered with a nucleotide donor template(s) (also referred to as "donor sequence(s)") with matching homology to the sites they are editing by utilizing the cell's DNA damage repair mechanism known as Homology Directed Repair (HDR). However, in practice the majority of mammalian cell types primarily effect DNA damage repair via NHEJ, which results in imprecise indels or nucleotide swaps. The various DNA damage pathways have various recruitment machinery that recognize the damage and begin the repairs and this machinery may be

manipulated to give more precise control over the repair. For NHEJ, the main recruitment proteins are DNA-dependent Protein Kinases (DNA-PK). It is contemplated by the present inventors that downregulating NHEJ will allow other DNA repair pathways, such as HDR, to be utilized, which will increase the accuracy of the repair or desired insertion.

[0010] Others have attempted to diminish NHEJ with other strategies but with limited success. Riesenberg, *et al.*, (US Patent Publication No. 2022/0154220) discloses inhibition of NHEJ using the DNA-PKcs (protein kinase catalytic subunit) inhibitor Nedesertib (M3814). However, the disclosure does not discuss or suggest the use of other PKcs inhibitors beside the spedific compound disclosed for improving CRISPR-based editing.

[0011] Riesenberg, *et al.*, (US Patent Publication No. 2020/0392540) discloses using mutated protein kinase subunits (among other compounds) to enhance DNA editing. However, the disclosure does not discuss or suggest the use of other non-mutated PK inhibitors alone outside of this mix for improving CRISPR-based editing.

[0012] Weinberg, *et al.*, (US Patent Publication No. 2021/0060028) discloses a specific Vertex Pharmaceuticals compound for the inhibition of DNA-dependent protein kinase. The disclosure is limited to the specific compound disclosed and for its use in cancer therapies.

[0013] Selvaraj, *et al.*, (WO2023/220418) discloses methods and compositions for editing DNA via the HDR pathway using the specific DNA-PK inhibitors taught in the specification.

[0014] Thus, what is also needed in the art are novel, inventive compositions and methods of both downregulating NHEJ in target cells to decrease indel formation and increase successful donor insertion and downregulating SSBR in addition to NHEJ in target cells to decrease indel formation and increase successful base conversation and gene editing and thereby increase the accuracy of the desired outcome.

Summary of the Invention

[0015] The present invention discloses compositions and methods for improvement of base editing and decrease in indel formation wherein cells are treated with inhibitors of DNA repair pathways. We demonstrate that inhibitors against non-homologous end joining (NHEJ) and single strand break repair (SSBR), when used alone or in combination, result in, for example, increased rates of C-to-T editing and reduced rates of undesired indel formation. Schematic summaries of these repair pathways can be found in Fig. 1. These inhibitors also provide benefit in the presence of an inhibitor of base excision repair, such as uracil glycosylase inhibitor, as well as in the presence of an RNP electroporation enhancer.

[0016] SSBR functions to repair single-strand breaks, such as the nick in the unedited strand caused by the nickase Cas effector (Fig. 2). This nick serves to mimic the single-strand breaks seen in Okazaki fragments during DNA replication. When the wrong nucleotide is incorporated into the newly synthesized strand during replication, the presence of these single-strand nicks signals to the mismatch repair machinery that the opposite strand should be used as the repair template; in base editing this nick results in the edited strand serving as the repair template and biases mismatch repair in the direction of installing the edit (Fig. 2, thick arrow). When SSBR is highly active, this nick is repaired and either strand may be used as the repair template, resulting in a lower rate of C-to-T editing. Inhibition of this pathway, for example by the PARP1/2 inhibitor Olaparib, preserves the nick in the unedited strand and results in an increase in the rate of C-to-T editing.

[0017] NHEJ is one of the default pathways for double-strand DNA break (DSB) repair in most mammalian cell types. However, it has a tendency to result in the installation of insertions and deletions (indels) at the site of the DSB. In the absence of NHEJ, other repair pathways, such as homology-directed repair (HDR), become more prevalent. In HDR, the homologous chromosome or sister chromatid is used as a template for the faithful repair of the DSB. In the setting of base editing,

this can result in installation of the edit (if the template DNA was edited) to generate a homozygous genetic change or a reversion to the unedited sequence (if the template DNA was unedited or faithfully repaired) (Fig. 3). In the latter case, the repaired DNA can be targeted by another base editor molecule for another opportunity at achieving a successful edit. Therefore, inhibition of NHEJ may simultaneously reduce the indel rate in edited cells and increase the rate of editing.

[0018] In summary, inhibition of NHEJ and/or SSBR DNA repair machinery, individually or simultaneously, serves to prevent the formation of undesired indels and increase the rate of installation of C-to-T substitutions. This results in a greater percentage of edited cells carrying the desired edit, resulting in reduced time, labor, and costs associated with screening of clones to identify the desired cell(s).

[0019] Further, inhibition of NHEJ serves to prevent the formation of undesired indels and increase the rate, accuracy and efficiency of desired insertions (*i.e.*, insertion of donor nucleotide sequences) because alternate repair pathways such as HDR are utilized by the target cell. This results in a greater percentage of edited cells carrying the desired insertion, resulting in reduced time, labor, and costs associated with screening of clones to identify the desired cell(s).

[0020] Thus, the present invention relates to methods and compositions suitable to enhance base editing by increasing the rate of base conversions and/or reducing the rate of indel formation, thereby increasing successful gene editing.

[0021] Further, the present invention relates to methods and compositions suitable for enhancing homology directed repair by increasing the rate of successful donor insertion and/or reducing the rate of indel formation, thereby increasing successful gene editing.

[0022] In one aspect, the present invention contemplates a method for improving the rate of nucleobase conversion, improving gene editing and/or reducing the rate of indel formation during CRISPR base editing, said method comprising: providing: i) a CRISPR base editor complex comprising: a) a catalytic effector domain, b) a catalytically modified Cas protein, and c) a guide RNA; to form the base editor

complex; ii) target cells; and iii) at least one inhibitor selected from one or more of single-strand break repair (SSBR) inhibitors, and nonhomologous end joining (NHEJ) inhibitors; introducing said base editor complex into target cells to form transfected cells, and; contacting said transfected cells with said one or more inhibitors of SSBR and/or NHEJ to form transfected and treated cells; wherein, said transfected and treated cells exhibit one or more of an increase in the rate of nucleobase conversion and/or a decrease in indel formation when compared to transfected but untreated cells.

[0023] In another aspect, the present invention further contemplates that the transfected cells are treated with both an inhibitor of SSBR and an inhibitor of NHEJ.

[0024] In another aspect, the present invention further contemplates that the transfected cells are treated with an inhibitor of SSBR.

[0025] In another aspect, the present invention further contemplates that the transfected cells are treated with an inhibitor of NHEJ.

[0026] In another aspect, the present invention further contemplates that the transfected and treated cells exhibit an increase in nucleobase conversion and where said nucleobase conversion is from 1 % to 200 % greater, 2% to 100% greater, 4% to 50% greater, 2% to 100% greater when compared to transfected but untreated cells or 10% to 60% greater when compared to transfected but untreated cells.

[0027] In another aspect, the present invention further contemplates that the transfected and treated cells exhibit an increase in nucleobase conversion and where said nucleobase conversion is at least 1 % greater, at least 5 % greater, at least 10 % greater, at least 20% greater, at least 30% greater, at least 40% greater, at least 50% greater, at least 60% greater, at least 70% greater, at least 70% greater, at least 80% greater, at least 90% greater, at least 100% greater, at least 150% greater, at least 200% greater, at least 500% greater, at least 1000% greater, or more, when compared to transcfected but untreated cells.

[0028] In another aspect, the present invention further contemplates that the transfected and treated cells exhibit a decrease in indel

formation and wherein said decrease in indel formation is at least 1%, at least 5%, at least 10% at least 20%, at least 30%, at least 40%, at least 50%, at least 60% at least 70%, at least 80% at least 90%, at least 95% or 100%, when compared to transfected but untreated cells.

[0029] In another aspect, the present invention contemplates that the decrease in indel formation is from about 2% to about 90 %, from about 5% to about 50% or about 50% to 90% when compared to transfeted but untreated cells.

[0030] In another aspect, the present invention contemplates that the increase in gene editing is from 2% to 100% or more, or 4% to 50%, 2% to 1000%, at least 1 % greater, at least 5 % greater, at least 10 % greater, at least 20% greater, at least 30% greater, at least 40% greater, at least 50% greater, at least 60% greater, at least 70% greater, at least 70% greater, at least 80% greater, at least 90% greater, at least 100% greater, at least 150% greater, at least 200% greater, at least 500% greater, at least 1000% greater, or more, when compared to transfected but untreated cells

[0031] In another aspect, the present invention contemplates that the nucleobase conversion is from cytosine (C) to thymine (T).

[0032] In another aspect, the present invention contemplates that the nucleobase conversion is from adenine (A) to guanine (G).

[0033] In another aspect, the present invention contemplates that the nucleobase conversion is from guanine (G) to cytosine (C) or thymine (T).

[0034] In another aspect, the present invention contemplates that the inhibitor of SSBR is a PARP inhibitor and any PARP inhibitor may be used in the methods of the present invention including, but not limited to, Olaparib (AZD2281), Veliparib (ABT-888) and A-966492.

[0035] In another aspect, the present invention contemplates that the inhibitor of NHEJ is an inhibitor of DNA-PK and any inhibitor of DNA-PK may be used in the methods of the present invention including, but not limited to, NU7441, AZD7648, CC-115 and M3814.

[0036] In another aspect, the present invention contemplates that the methods of the present invention may additionally comprise a uracil

glycosylase inhibitor (UGI) or a nucleotide sequence encoding UGI in the transfection mixture.

[0037] In another aspect, the present invention contemplates that dextran sulfate is included in the transfection mixture. It is still further contemplated that the concentration of dextran sulfate is from 0.1 µg/ml to 10 µg/ml, approximately 1.0 µg/ml or less than 10 µg/ml.

[0038] In another aspect, it is contemplated in the present invention that the CRISPR base editing complex is encoded in one or more nucleic acid sequences.

[0039] In another aspect, it is contemplated in the present invention that said SSBR inhibitor is the PARP1/2 inhibitor AZD2281 (Olaparib).

[0040] In another aspect, it is contemplated that in the present invention the SSBR PARP1/2 inhibitor is at a concentration of, for example, 0.1 µM to 10 µM, 0.2 µM to 8.0 µM, 0.5 µM to 5.0 µM, less than 10 µM, less than 8.0 µM, less than 5.0 µM,less than 2.0 µM or less than 0.5 µM. One of skill in the art may determine, aided with the guidance provided by this specification, that for a specific use the SSBR PARP1/2 inhibitor may be used a concentration different than the range stated above.

[0041] In another aspect, it is contemplated that in the present invention the NHEJ inhibitor are the DNA-dependent protein kinase inhibitors MSC2528072A and MSC2528070A (*see*, Fig. 16 and US 9,732,094, which is incorporated herein by reference: Merck Health Care, Darmstadt, DE).

[0042] In another aspect, it is contemplated that the present invention the NHEJ inhibitor is at a concentration of 0.1 µM to 10 µM although one of skill in the art may determine, aided with the guidance provided by this specification, that for a specific use the NHEJ inhibitor may be used a concentration different than the range stated above. For example, the concentration used may be 0.2 µM to 8.0 µM, 0.5 µM to 5.0 µM, less than 10 µM, less than 8.0 µM, less than 5.0 µM,less than 2.0 µM or less than 0.5 µM.

[0043] It is further contemplated that the when the transfected and treated cells of the present invention are compared to transfected but

untreated cells, the comparison is based on historic data of analogous transfected and untreated cells. It is still further contemplated that the rate of nucleobase conversion in transfected and treated cells is increased when compared to analogous transfected and untreated cells. It is still further contemplated that the rate of indel formation in transfected and treated cells is decreased when compared to analogous transfected and untreated cells.

[0044] It is still further contemplated that the base editing of the present invention is performed in a eukaryotic cell, wherein said eukaryotic cell is *in vivo*, and, additionally is performed on one or more somatic cells or in a non-human embryo.

[0045] It is further contemplated that the Cas protein is a catalytically modified and the catalytically modified Cas protein is a nickase and said nickase generates a single strand break on the targeted strand.

[0046] It is further contemplated that the Cas protein is a catalytically modified Cas protein and the catalytically modified Cas protein is catalytically inactive.

[0047] The present invention further contemplates that the base editor complex is introduced into target cells to form transfected cells by any of electroporation, microinjection, lipofection, calcium phosphate-mediated transfection, nucleofection, cationic polymer transfection, viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection and magnetofection.

[0048] The present invention further contemplates a method for reducing the rate of indel formation and/or increasing gene editing by homology-directed repair (HDR), said method comprising: providing: i) a site-specific or RNA-guided endonuclease; ii) a donor nucleotide sequence, iii) target cells; and iv) at least one nonhomologous end joining (NHEJ) inhibitor; introducing said nuclease and donor nucleotide sequence into target cells to form transfected cells; and; contacting said transfected cells with said one or more inhibitors of NHEJ to form transfected and treated cells; wherein, said transfected

and treated cells exhibit a decrease in indel formation and/or an increase in HDR when compared to transfected but untreated cells.

[0049] The present invention further contemplates that the the RNA-guided endonuclease is selected from a zinc-finger nuclease, a TALEN, a meganuclease, Fanzor nuclease, or a CRISPR complex comprising a Cas protein and a guide RNA.

[0050] The present invention still further contemplates that the transfected and treated cells exhibit an increase in HDR when compared to transfected but untreated cells.

[0051] The present invention still further contemplates that the transfected and treated cells exhibit a decrease in indel formation when compared to transfected but untreated cells.

[0052] The present invention further contemplates that the transfected and treated cells exhibit an increase in HDR that is at least 1% or greater, at least 5% or greater, at least 10% or greater, at least 20% or greater, at least 30% or greater, at least 40% or greater, at least 50% or greater, at least 60% or greater, at least 70% or greater, at least 80% or greater, at least 90% or greater, at least 100% or greater, at least 150% or greater, at least 200%or greater, at least 500% or greater, at least 1000% or greater, or more, when compared to untreated cells.

[0053] The present invention further contemplates that the transfected and treated cells exhibit an increase in HDR that is from at least 2%, or 2% to 90% or from 5% to 50%, or from 2% to 100% or from 2% to 1000% when compared to transfected but untreated cells

[0054] The present invention further contemplates that the transfected and treated cells exhibit a decrease in indel formation and wherein said decrease in indel formation is at least 1%, at least 5%, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99.9% or 100%, when compared to untreated cells.

[0055] The present invention further contemplates that the transfected and treated cells exhibit a decrease in indel formation from at least 2%,

or 2% to 90%, or from 5% to 50%, or from 2% to 100% when compared to transfected but untreated cells.

[0056] The present invention further contemplates that the inhibitor of NHEJ is an inhibitor of DNA-PK and is selected from a group consisting of NU7441, AZD7648, CC-115 and M3814.

[0057] The present invention further contemplates that the NHEJ inhibitor is the DNA-dependent protein kinase inhibitor MSC2528072A and/or MSC2528070A.

[0058] The present invention further contemplates that the NHEJ inhibitor is at a concentration of 0.1 μM to 10 μM, contemplated that in the present invention the NHEJ inhibitor is at a concentration of, for example, 0.1 μM to 10 μM, 0.2 μM to 8.0 μM, 0.5 μM to 5.0 μM, less than 10 μM, less than 8.0 μM, less than 5.0 μM, less than 2.0 μM or less than 0.5 μM or greater than 0.1 μM. One of skill in the art may determine, aided with the guidance provided by this specification, that for a specific use the NHEJ inhibitor may be used a concentration different than the range stated above.

[0059] The present invention further contemplates that the comparison of transfected and treated cells to transfected but untreated cells is based on historic data of analogous transfected but untreated cells.

[0060] The present invention further contemplates that the endonuclease-based editing is performed in a eukaryotic cell.

[0061] The present invention further contemplates that the endonuclease-based editing method is performed *in vivo*.

[0062] The present invention further contemplates that the endonuclease-based editing method is performed *in vitro*.

[0063] The present invention further contemplates that the endonuclease-based editing method is performed in one or more somatic cells.

[0064] The present invention further contemplates that the endonuclease-based editing method is performed in non-human embryos.

[0065] The present invention further contemplates that the endonuclease has full endonuclease activity.

[0066] The present invention further contemplates that the endonuclease is catalytically modified, that in one embodiment the catalytically modified endonuclease is a nickase that generates a single strand break on the targeted strand or is enzymatically inactive.

[0067] The present invention further contemplates that the endonuclease is introduced into target cells to form transfected cells by a method selected from the group consisting of: electroporation, microinjection, lipofection, calcium phosphate-mediated transfection, nucleofection, cationic polymer transfection, viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection and magnetofection.

[0068] The present invention further contemplates that the endonuclease is encoded in one or more nucleic acid sequences and, in some embodiments, said nuclease sequences are transfected into target cells.

Brief Description of the Drawings

[0069] Figs. 1A & 1B show schematic summaries of SSBR and NHEJ repair pathways.

[0070] Fig. 2 shows a schematic summary of the role of a single-strand nick and the effects of single strand break repair in base editing.

[0071] Fig. 3 shows a schematic summary of a double strand break repair in base editing.

[0072] Figs. 4A & 4B show rates of indel formation during base editing. A) Target EMX1-11. B) Target HBB.

[0073] Fig. 5 shows rates of indel formation during base editing when UGI and/or dextran sulfate (DS) are present during transfection.

[0074] Fig. 6 shows SEQ ID NO: 1, APOBEC3B-SpCas9 nickase nucleotide sequence) and SEQ ID NO: 2, UGI with nuclear location sequence (NLS) peptide sequence (Italic = linker; underline = NLS).

[0075] Fig. 7 shows SEQ ID NOs: 21 – 28. These are the guide spacer, ssODN (single-stranded oligodeoxynucleotides) and next generation sequencing (NGS) primer sequences used in Example 1.

[0076] Fig. 8 shows toxicity analysis results with CellTiterGLO v2 for the BCL4 target.

[0077] Fig. 9 shows the raw numbers used to generate Fig. 8.

[0078] Fig. 10 shows toxicity analysis results with CellTiterGLO v2 for the AAVS1 target.

[0079] Fig. 11 shows the raw numbers used to generate Fig. 10.

[0080] Fig. 12 shows NGS analysis for AAVS1 in K562 cells.

[0081] Fig. 13 shows the raw numbers used to generate Fig. 12.

[0082] Fig. 14 shows NGS analysis for BCL-4 in K562 cells.

[0083] Fig. 15 shows the raw numbers used to generate Fig. 14.

[0084] Fig. 16 shows the molecular structures of MSC2528070A and MSC2528072A.

Detailed Description of the Invention

[0085] **Definitions**

[0086] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton, *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger, *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0087] When introducing elements of the present disclosure or the preferred embodiments(s) thereof, the articles "a," "an," "the" and "said" are intended to mean that there are one or more of the elements. The terms "comprising," "including" and "having" are intended to be

inclusive and mean that there may be additional elements other than the listed elements.

[0088] Furthermore, the transitional phrases “comprising,” “consisting essentially of” and “consisting of” have the meanings as given in MPEP 2111.03 (Manual of Patent Examining Procedure; United States Patent and Trademark Office, 9th Ed., Revision Feb 2023 [R-07.2022]). Any claims using the transitional phrase “consisting essentially of” will be understood as reciting only essential elements (*i.e.*, the basic and novel characteristics) of the invention and any other elements recited in dependent claims are understood to be non-essential to the invention recited in the claim from which they depend. Likewise, any additional elements over those claimed that are described in a prior art reference(s) are excluded from the claims by use of the transitional phrase “consisting essentially of.”

[0089] As used herein, the term "endogenous sequence" refers to a chromosomal sequence that is native to the cell.

[0090] The term "exogenous" or “exogenous sequence,” as used herein, refers to a sequence that is not native to the cell or a chromosomal sequence whose native location in the genome of the cell is in a different chromosomal location.

[0091] A "gene," as used herein, refers to a DNA region (including exons and introns) encoding a gene product, as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites, and locus control regions.

[0092] The terms “complementary” or “complementarity” refer to the association of double-stranded nucleic acids by base pairing through specific hydrogen bonds. The base pairing may be standard Watson-Crick base pairing (*e.g.*, 5′-A G T C-3′ pairs with the complementary

sequence 3′-T C A G-5′). The base pairing also may be Hoogsteen or reversed Hoogsteen hydrogen bonding. Complementarity is typically measured with respect to a duplex region and, thus, excludes overhangs, for example. Complementarity between two strands of the duplex region may be partial and expressed as a percentage (*e.g.*, 70%), if only some of the base pairs are complementary. The bases that are not complementary are "mismatched." Complementarity may also be complete (*i.e.*, 100%), if all the base pairs of the duplex region are complementary.

[0093] The term "homologous" refers to the extent two or more sequences are identical. Two sequences are considered to be homologous if they will hybridize to the same sequence under a defined set of conditions. Defined conditions include but are not limited to buffer formulation, temperature and sequence concentration.

[0094] The term "heterologous" refers to an entity that is not endogenous or native to the cell of interest. For example, a heterologous protein refers to a protein that is derived from or was originally derived from an exogenous source, such as an exogenously introduced nucleic acid sequence. In some instances, the heterologous protein is a protein not normally produced by the cell of interest.

[0095] The terms "nucleic acid" and "polynucleotide" refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogs of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (*e.g.*, phosphorothioate backbones). In general, an analog of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analog of A will base-pair with T.

[0096] The term "synthetic nucleic acid" refers to a nucleotide sequence synthesized *in vitro* (for example, in a lab and either manually or with a nucleic acid synthesizer device) and in which the sequence is not found in nature. The sequence may be, for example,

DNA or RNA or a modification thereof as described below, may be any length and may be any sequence of nucleotides so long as the sequence is not naturally occurring.

[0097] The term "nucleotide" refers to deoxyribonucleotides or ribonucleotides. The nucleotides may be standard nucleotides (*i.e.*, adenosine, guanosine, cytidine, thymidine, and uridine) or nucleotide analogs. A nucleotide analog refers to a nucleotide having a modified purine or pyrimidine base or a modified ribose moiety. A nucleotide analog may be a naturally occurring nucleotide (*e.g.*, inosine) or a non-naturally occurring nucleotide. Non-limiting examples of modifications on the sugar or base moieties of a nucleotide include the addition (or removal) of acetyl groups, amino groups, carboxyl groups, carboxymethyl groups, hydroxyl groups, methyl groups, phosphoryl groups, and thiol groups, as well as the substitution of the carbon and nitrogen atoms of the bases with other atoms (*e.g.*, 7-deaza purines). Nucleotide analogs also include dideoxy nucleotides, 2'-O-methyl nucleotides, locked nucleic acids (LNA), peptide nucleic acids (PNA), and morpholinos.

[0098] The terms "polypeptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues.

[0099] The term “inactivated,” in the context of the present invention, may refer to the deletion of one or more amino acids and/or the substitution of one or more amino acids or the addition of one and/or more amino acids in a target protein such that one or more functions of the protein are eliminated or reduced to a level where activity is less than 95%, less than 90%, less than 75%, less than 50%, less than 40%, less than 30%, less than 20%, less than 15%, less than 10%, less than 5%, less than 4%, less than 3%, less than 2% or less than 1% of the level of activity of the active protein. In the present invention, the Cas-like protein (*e.g.*, Cas9) is catalytically inactivated as the result of having two amino acids substituted with Alanine residues as detailed elsewhere in this specification thereby inhibiting its ability to cleave double-stranded DNA. Reduced activity may be referred to “partially inactivated.” The term “catalytically modified” in the context of the

present invention refers to a subject protein that is inactivated or partially inactivated and means that one or more of the catalytic activities of the subject protein have been reduced or eliminated. The nuclease activity of the Cas protein of the present invention may be partially or fully inactivated making the Cas protein a nickase (cleaving only one DNA strand; discussed in greater detail, below) or catalytically inactive (cleaving none of the DNA strands), respectively.

[0100] The term "directly linked" with regard to proteins and polypeptides in the context of the present invention means that two proteins are joined together (*i.e.*, fused, *e.g.*, via peptide bonds) to form a continuous protein or polypeptide with no added amino acid residues incorporated between the two joined/fused proteins.

[0101] The term "indirectly linked" in the context of the present invention means that one or more amino acids are incorporated between two joined/fused proteins.

[0102] The terms "inhibit" and "inhibition" with regard to the present invention, shall mean total or partial inhibition. Thus, inhibition can be from over 0% to 100%, from 1% to 100%, from 5% to 100%, from 10% to 100%, from 20% to 100%, from 30% to 100%, from 40% to 100%, from 50% to 100%, from 60% to 100%, from 70% to 100%, from 80% to100 %, from 90% to 100%, from 95% to 100% and from 99% to 100%. The lever of inhibition can be over 0%, over 1%, over 5%, over 10%, over 20%, over 30%, over 40% over 50%, over 60%, over 70% over 80%, over 90%, over 95% and over 99%. The level of inhibition can be 100% and under but not less than 0%. In the present invention, indel formation is inhibited during CRISPR-based gene editing. In the present invention it is contemplated that inhibitors of NHEJ are added to cell cultures of the target cells before transfection, during transfection or after transfection of the cells, wherein the transfection is used to insert the CRISPR gene editing complex and donor sequence into the target cells.

[0103] Techniques for determining nucleic acid and amino acid sequence identity are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene

and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion. In general, “identity” refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their “percent identity." The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, Wis.) in the "BestFit" utility application. Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art. For example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Details of these programs can be found on the GenBank website.

[0104] “Inhibitor of PARP” or “PARP inhibitor” refers to a reagent or reagents that block, partially or totally, PARP function and/or the PARP pathway. Although the present invention is not limited by theory, it is

reported that PARP (Poly(ADP-ribose) polymerase) 1 (PARP1), also known as ADPRT1, is a multifunctional human ADP-ribosyltransferase. It plays a role in multiple DNA repair pathways, including the base excision repair (BER), non-homologous end joining (NHEJ), homologous recombination (HR), and Okazaki-fragment processing pathways. In response to DNA strand breaks, PARP1 covalently attaches ADP-ribose moieties to arginine, glutamate, aspartate, cysteine, lysine, and serine acceptor sites on both itself and other proteins. This signal recruits DNA repair proteins to the site of DNA damage. PARP1 binding to these sites enhances ADP-ribosylation via allosteric communication between the distant DNA binding and catalytic domains. *See*, Spiegel JO, Van Houten B, Durrant JD. PARP1: Structural insights and pharmacological targets for inhibition. DNA Repair (Amst). 2021 Jul;103:103125. doi: 10.1016/j.dnarep.2021.103125. Epub 2021 Apr 14. PMID: 33940558; PMCID: PMC8206044, which is representative of the knowledge available to one of skill in the art at the time of the invention.

[0105] PARP inhibitors are known to one of skill in the art and may be used in the methods taught herein in view of the teaching of this specification. Non-limiting examples of known PARP inhibitors include: pamiparib, veliparib, olaparib, rucaparib, E7449, niraparib, and talazoparib. *See*, *e.g.*, Petar-Bogomil Kanev, Sylvia Varhoshkova, Irina Georgieva,et al. A unified mechanism for PARP inhibitor-induced PARP1 chromatin retention at DNA damage sites in living cells, Cell Reports, Volume 43, Issue 5, 2024, 114234, ISSN 2211-1247, https://doi.org/10.1016/j.celrep.2024.114234. Suitable concentrations for use in the present invention range from approximately 0.1 μM to 10 μM but may also range higher or lower. One of skill in the art will be able to determine suitable concentrations for a particular use with the guidance provided by this specification. Other suitable concentrations range from 0.2 μM to 9 μM, 0.3 μM to 8 μM, 0.4 μM to 7 μM,d 0.5 μM to 6 μM, 5.0 μM to 10.0 μM and 1.0 μM to 0.9 μM. Further, suitable concentrations may be any concentration 0.01 μM or higher, 0.05 μM or higher, 0.1 μM or higher, 0.5 μM or higher, 1.0 μM or higher, 2.0 μM

or higher, 5.0 μM or higher and up to 10 μM or higher, providing that the dose is not toxic or detrimental to the cells being treated.

[0106] "Inhibitor of DNA-PK" or "DNA-PK inhibitor" refers to reagents that block, partially or totally, DNA-PK (DNA-protein kinase) function or pathway. Although the present invention is not limited by theory, it is reported that DNA-PK (a member of the phosphatidylinositol 3-kinase-related kinases family – PIKK) plays an important role in the detection and repair of DNA double-strand breaks via the non-homologous end-joining pathway. *See*, Pospisilova M, Seifrtova M, Rezacova M. Small molecule inhibitors of DNA-PK for tumor sensitization to anticancer therapy. J Physiol Pharmacol. 2017 Jun;68(3):337-344. PMID: 28820390, which is representative of the knowledge available to one of skill in the art at the time of the invention.

[0107] DNA-PK inhibitors are known to one of skill in the art and may be used in the methods taught herein in view of the teaching of this specification. One non-limiting example of a known DNA-PK inhibitor is: AZD-7648. Others are given in the paragraphs directly above and below. Suitable concentrations for use in the present invention range for approximately 0.1 μM to 10 μM but may also range higher or lower. One of skill in the art will be able to determine suitable concentrations with the guidance provided by this specification. Other suitable concentrations range from 0.2 μM to 9 μM, 0.3 μM to 8 μM, 0.4 μM to 7 μM,0.5 μM to 6 μM, 5.0 μM to 10.0 μM and 1.0 μM to 0.9 μM. Further, suitable concentrations may be any concentration 0.01 μM or higher, 0.05 μM or higher, 0.1 μM or higher, 0.5 μM or higher, 1.0 μM or higher, 2.0 μM or higher, 5.0 μM or higher and up to 10 μM or higher, providing that the dose is not toxic or detrimental to the cells being treated.

[0108] Further still, an inhibitor of DNA-PK may refer to any and all compounds disclosed in US Patent Publication No. 2022/0117970 (the '970 publication) to Merck Patent GmbH and any parent applications and issued patents from which the '970 publication cites priority including but not limited to WO 2014/183850, US 9,732,094, US 10,172,859, US 10,383,874, US 11,065,253 and any and all foreign

counterparts, all of which are incorporated herein by refence in their entirety. In an embodiment, it is contemplated that the PKi inhibitors are selected from one or more of MSC2528072A and/or MSC2528070A (Merck KGaA, Darmstadt, DE) (*see*, Fig. 16 and US 9,732,094, which is incorporated herein by reference).

[0109] **CRISPR/Cas Proteins and Systems**

[0110] Understanding the present invention will be aided by understanding CRISPR/Cas protein systems in general and in the context of the present invention.

[0111] **The CRISPR/Cas9-based System**

[0112] In its basic form, the CRISPR/Cas9 system introduces a double strand break close to the binding site of a guide RNA (gRNA). A Cas9 protein and a gRNA form a ribonucleoprotein (RNP) complex. This complex itself can be transfected into a recipient cell (for example, via the use of extracellular vesicles) or a plasmid or viral vector encoding the complex can be used. The gRNA targets the complex to the desired location in the genome through formation of an R-loop – a three-stranded nucleic acid structure composed of a DNA:RNA hybrid and a displaced single-stranded DNA – where the Cas9 protein (an endonuclease) causes a double-strand DNA break where modifications to the sequence can be made. Such modifications can take the form of non-homologous end joining (NHEJ) making random small insertions or deletions (indels) or homology-directed repair (HDR). NHEJ is useful for creating knockout mutations and HDR is useful for making desired modifications to the target sequence, also termed "precision editing." While HDR enables the introduction of these precision edits, its use without DNA-PK inhibition is often hindered by low rates of installation of the desired edit and high rates of indels.

[0113] **Base Editors**

[0114] As discussed above, precision editing using traditional CRISPR/Cas9 technology has challenges regarding efficiency and a

high rate of undesired indels. Base editing was developed (Komor, 2016) to increase the efficiency of precision editing using the CRISPR/Cas9 technology while minimizing the rate of indels. Base editing allows for the direct, irreversible conversion of one target DNA base pair into another without requiring a dsDNA backbone break or a donor template. Rather, a target nucleobase is converted into another by a catalytic effector domain tethered to the RNP complex. For cytosine base editing, the tethered catalytic domain is a cytosine deaminase, which chemically converts a cytosine nucleobase (C) into a uracil (U). Uracil has the base-paring properties of thymine (T). Thus, upon DNA replication or repair, the targeted C:G pair is converted to a T:A pair. For adenine base editing, the tethered catalytic domain is an adenine deaminase, which chemically converts adenine (A) into inosine (I). Inosine has the base-pairing properties of guanine (G). Thus, upon DNA replication or repair, the targeted A:T pair is converted to a G:C pair. For guanine base editing, the tethered catalytic domain is a guanine-specific glycosylase, which excises the guanine nucleobase from the DNA. Translesion synthesis at the abasic site leads to the installation of C or T, resulting in the installation of a G:C to C:G or T:A substitution.

[0115] **Site-Specific Endonucleases**

[0116] Methods other than RNA-guided endonucleases are known to generate double strand DNA breaks. These include but are not limited to zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and meganucleases. In general, ZFNs are restriction enzymes comprised of a fused zinc finger DNA-binding domain to a DNA-cleavage domain, made to target specific DNA sequences. The ZFN DNA-binding domain is made of between 3 to 6 zinc finger repeats that can each target from 9 to 18 basepairs, allowing for unique and targeted DNA binding. The ZFN DNA-cleavage domain from the FokI endonuclease is nonspecific and must dimerize in order to cleave DNA. To enable dimerization, two individual ZFNs

must bind opposite strands of DNA with their C-termini at a specific distance.

[0117] TALENs are also restriction enzymes that are engineered to cut a targeted DNA sequence, made by a fused TAL effector DNA-binding domain to a DNA cleavage domain. This DNA-binding domain has two amino acids in its sequence that are variable and linked to recognition with specific nucleotides, and engineering these two amino acids allows for unique targeted DNA binding. As in ZFNs, TALENS utilize the FokI endonuclease for DNA cleavage, which requires dimerization . As a result, TALENs require two constructs with unique DNA-binding domains with proper orientation and spacing to achieve this dimerization.

Meganucleases are a group of endodeoxyribonucleases that have a large recognition site of 12 to 40 basepairs, making them highly specific restriction enzymes. The high precision of these meganucleases limits their usability for targeted DNA-cleavage. Unique, targeted DNA-binding meganucleases can be engineered by modifying the amino acid sequence that determines specificity or by fusing protein domains from different enzymes together.

A scientist sufficiently skilled in the art should know how to use these to effect HDR, as each creates a double stranded break in the chromosomal sequence.

[0118] **RNA-Guided Endonucleases**

[0119] RNA-guided endonucleases, such as Cas9, may comprise at least one nuclear localization signal (NLS), at least one nuclease domain, and at least one domain that interacts with a guide RNA (gRNA) to direct the endonuclease/deaminase complex of the present invention to a specific nucleobase for modification. Also known are nucleic acids encoding the RNA-guided endonucleases, as well as methods of using the RNA-guided endonucleases to modify chromosomal sequences of eukaryotic cells or embryos. The RNA-guided endonuclease interacts with specific gRNAs, each of which directs the endonuclease to a specific targeted site, at which site the

catalytic effector domain modifies the target nucleobase, resulting in the installation of a substitution. Since the specificity is provided by the gRNA, the RNA-guided endonuclease is, essentially, universal and can be used with different gRNAs to target different genomic sequences. The methods disclosed herein can be used to target and modify specific chromosomal sequences at targeted locations in the genome of cells or embryos. Furthermore, the targeting is specific with limited off-target effects.

[0120] Many forms of guide RNAs (gRNAs) are known in the field. In general, a gRNA is an RNA molecule that can direct an RNA binding protein or endonuclease to a specific nucleic acid sequence/target site by means of base pairing. A gRNA may comprise a single RNA molecule, such as a crisprRNA (crRNA), or two RNA molecules, such as a crRNA and a tracrRNA (trRNA). In some embodiments, it may additionally comprise an accessory RNA or DNA molecule. In some embodiments, a crRNA and a trRNA may be covalently linked together to form a chimeric guide RNA or a single guide RNA (sgRNA). It is also well established in the field that a gRNA can be introduced alone or in combination with its cognate gRNA binding protein or endonuclease into a target cell in different forms. It can be expressed from a DNA vector introduced into a target cell. It can be synthesized by *in vitro* transcription or by chemical reaction before being introduced into a target cell. A person of skill in the art should know that many chemical modifications are possible on a gRNA by chemical synthesis. For example, certain modifications, such as 2'-O methyl group modification and phosphorothioate linkage modification, may be introduced into a gRNA to alter its stability or performance. Nonstandard nucleotides, such as DNA and LNA, also may be introduced into a gRNA during chemical synthesis to alter its specificity or performance. A guide RNA that may have a different form or may contain a different chemical modification may be used in conjunction with the present disclosure without deviating from the spirit of the present disclosure.

[0121] The present disclosure provides fusion proteins, wherein a fusion protein comprises a CRISPR/Cas-like protein or fragment

thereof wherein the Cas-like protein is fully or partially catalytically inactivated, retaining its ability to bind DNA but without being able to generate double-strand breaks in the target DNA. Each fusion protein is guided to a specific chromosomal sequence by a specific gRNA, wherein the associated (*e.g.*, tethered or otherwise linked) catalytic effector domain converts the target nucleobase to another nucleobase.

[0122] RNA-guided endonucleases may comprise at least one nuclear localization signal (NLS), which permits entry of the endonuclease into the nuclei of eukaryotic cells and embryos such as, for example, non-human one cell embryos. RNA-guided endonucleases also comprise at least one nuclease domain and at least one domain that interacts with a gRNA. An RNA-guided endonuclease is directed to a specific nucleic acid sequence (or target site) by a gRNA. The gRNA interacts with the RNA-guided endonuclease as well as the target site such that, once directed to the target site, the associated catalytic effector domain can convert the target nucleobase to another nucleobase, thereby installing a substitution. Since the gRNA provides the specificity for the targeted cleavage the RNA-guided endonuclease is universal (providing that its ability to cleave DNA is disabled) and can be used with different gRNAs to modify different target residues. RNA-guided endonucleases can be proteins, can be encoded by isolated nucleic acids (*i.e.*, RNA or DNA), can be encoded by vectors comprising nucleic acids encoding the RNA-guided endonucleases, and can be protein-RNA complexes comprising the RNA-guided endonuclease plus a gRNA.

[0123] The RNA-guided endonuclease can be derived from a clustered regularly interspersed short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system. The CRISPR/Cas system can be a type I, a type II, or a type III system, as known to one of ordinary skill in the art. Non-limiting examples of suitable CRISPR/Cas proteins include Cas3, Cas4, Cas5, Cas5e (or CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, Cas12, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (or CasA), Cse2 (or CasB), Cse3 (or CasE), Cse4 (or CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2,

Csb3,Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4, and Cu1966. One of ordinary skill in the art will be able to modify any RNA-guided nuclease to inactivate its catalytic activity for use with a base editor system.

[0124] In one embodiment, the RNA-guided endonuclease is derived from a type II CRISPR/Cas system. In specific embodiments, the RNA-guided endonuclease is derived from a Cas9 protein. The Cas9 protein can be from, for example, *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus sp.*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocosphaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor becscii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosipho africanus*, or *Acaryochloris marina*.

[0125] In general, CRISPR/Cas proteins comprise at least one RNA recognition domain and/or RNA binding domain. RNA recognition and/or RNA binding domains interact with guide RNAs. CRISPR/Cas proteins also usually comprise nuclease domains (*i.e.*, DNase or RNase domains; but for the purposes of base editors one or more of these are disabled), DNA binding domains, helicase domains, RNase

domains, protein-protein interaction domains, dimerization domains, as well as other domains.

[0126] The CRISPR/Cas-like protein can be a wild type CRISPR/Cas protein, a modified CRISPR/Cas protein, or a fragment of a wild type or modified CRISPR/Cas protein. The CRISPR/Cas-like protein can be modified to increase nucleic acid binding affinity and/or specificity, alter an enzymatic activity (*e.g.*, inactivate its ability to cleave DNA), and/or change another property of the protein. For example, nuclease (*i.e.*, DNase, RNase) domains of the CRISPR/Cas-like protein can be modified, deleted, or inactivated. Alternatively, the CRISPR/Cas-like protein can be truncated to remove domains that are not essential for the function of the fusion protein. The CRISPR/Cas-like protein can also be truncated or modified to optimize the activity of the effector domain of the fusion protein.

[0127] In some embodiments, the CRISPR/Cas-like protein can be derived from a wild type Cas9 protein or fragment thereof. In other embodiments, the CRISPR/Cas-like protein can be derived from a modified Cas9 protein. For example, the amino acid sequence of the Cas9 protein can be modified to alter one or more properties (*e.g.*, nuclease activity, affinity, stability, etc.) of the protein. Alternatively, domains of the Cas9 protein not involved in RNA-guided cleavage can be eliminated from the protein such that the modified Cas9 protein is smaller than the wild type Cas9 protein.

[0128] In general, a Cas9 protein comprises at least two nuclease (*i.e.*, DNase) domains. For example, a Cas9 protein can comprise a RuvC-like nuclease domain and an HNH-like nuclease domain. The RuvC and HNH domains work together to cut single strands to make a double-stranded break in DNA (Jinek, *et al.*, Science, 337: 816-821). In some embodiments, the Cas9-derived protein can be modified to contain only one functional nuclease domain (either a RuvC-like or a HNH-like nuclease domain). For example, the Cas9-derived protein can be modified such that one of the nuclease domains is deleted or mutated such that it is no longer functional (*i.e.*, the nuclease activity is absent). In some embodiments in which one of the nuclease domains

is inactive, the Cas9-derived protein is able to introduce a nick into a double-stranded nucleic acid (such protein is termed a "nickase"), but not cleave both strands of the double-stranded DNA. For example, an aspartate to alanine (D10A) conversion in a RuvC-like domain converts the Cas9-derived protein into a nickase. Likewise, a histidine to alanine (H840A or H839A) conversion in a HNH domain converts the Cas9-derived protein into a nickase. Each nuclease domain can be modified using well-known methods, such as site-directed mutagenesis, PCR-mediated mutagenesis, and total gene synthesis, as well as other methods known in the art. Modifying both of these domains results in inactivation of nuclease and nickase activity.

[0129] The RNA-guided endonuclease may comprise at least one NLS. In general, an NLS comprises a stretch of basic amino acids. Nuclear localization signals are known in the art (see, *e.g.*, Lange, *et al.*, J. Biol. Chem., 2007, 282:5101-5105). For example, in one embodiment, the NLS can be a monopartite sequence, such as PKKKRKV (SEQ ID NO: 9) or PKKKRRV (SEQ ID NO: 10). In another embodiment, the NLS can be a bipartite sequence. In still another embodiment, the NLS can be KRPAATKKAGQAKKKK (SEQ ID NO: 11). The NLS can be located at the N-terminus, the C-terminus, or in an internal location of the RNA-guided endonuclease.

[0130] In some embodiments, the RNA-guided endonuclease can further comprise at least one cell-penetrating domain. In one embodiment, the cell-penetrating domain can be a cell-penetrating peptide sequence derived from the HIV-1 TAT protein. As an example, the TAT cell-penetrating sequence can be GRKKRRQRRRPPQPKKKRKV (SEQ ID NO: 12). In another embodiment, the cell-penetrating domain can be TLM (PLSSIFSRIGDPPKKKRKV; SEQ ID NO: 13), a cell-penetrating peptide sequence derived from the human hepatitis B virus. In still another embodiment, the cell-penetrating domain can be MPG (GALFLGWLGAAGSTMGAPKKKRKV; SEQ ID NO: 14 or GALFLGFLGAAGSTMGAWSQPKKKRKV; SEQ ID NO: 15). In an additional embodiment, the cell-penetrating domain can be Pep-1

(KETWWETWWTEWSQPKKKRKV; SEQ ID NO: 16), VP22, a cell penetrating peptide from Herpes simplex virus, or a polyarginine peptide sequence. The cell-penetrating domain can be located at the N-terminus, the C-terminus, or in an internal location of the protein.

[0131] In still other embodiments, the RNA-guided endonuclease can also comprise at least one marker domain. Non-limiting examples of marker domains include fluorescent proteins, purification tags, and epitope tags. In some embodiments, the marker domain can be a fluorescent protein. Non limiting examples of suitable fluorescent proteins include green fluorescent proteins (*e.g.*, GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (*e.g.*, YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1,), blue fluorescent proteins (*e.g.,* EBFP, EBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire,), cyan fluorescent proteins (*e.g.*, ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRasberry, mStrawberry, Jred), and orange fluorescent proteins (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato) or any other suitable fluorescent protein. In other embodiments, the marker domain can be a purification tag and/or an epitope tag. Exemplary tags include, but are not limited to, glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly(NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis (SEQ ID NO: 17), biotin carboxyl carrier protein (BCCP), and calmodulin.

[0132] In certain embodiments, the RNA-guided endonuclease may be part of a protein-RNA complex comprising a gRNA. The gRNA interacts with the RNA-guided endonuclease to direct the endonuclease to a specific target site, wherein, for example, the 5' end of the guide RNA base pairs with a specific protospacer sequence.

[0133] **Fusion Proteins**

[0134] In some embodiments, an aspect of the present disclosure provides a fusion protein comprising a CRISPR/Cas-like protein or fragment thereof, a catalytic effector domain, and, for example, one or more UGI. The CRISPR/Cas-like protein is directed to a target site by a gRNA, at which site the effector modifies or affects the targeted nucleic acid sequence. In the present invention the "effector domain" can be a linked deaminase or glycosylase. The fusion protein can further comprise at least one additional domain chosen from a nuclear localization signal, a cell-penetrating domain, or a marker domain.

[0135] **CRISPR/Cas-Like Protein**

[0136] The fusion protein comprises a CRISPR/Cas-like protein or a fragment thereof. CRISPR/Cas-like proteins are detailed above in section (I). The CRISPR/Cas-like protein can be located at the N-terminus, the C-terminus, or in an internal location of the fusion protein.

[0137] In some embodiments, the CRISPR/Cas-like protein of the fusion protein can be derived from a Cas9 protein. Cas9-derived proteins can be wild type, modified, or a fragment thereof. In some embodiments of the present invention, the Cas9-derived protein is modified to inactivate both functional nuclease domains (both a RuvC-like and an HNH-like nuclease domain). In some embodiments, both of the RuvC-like nuclease domain and the HNH-like nuclease domain can be modified or eliminated such that the Cas9-derived protein is unable to nick or cleave double stranded nucleic acid. In still other embodiments, all nuclease domains of the Cas9-derived protein can be modified or eliminated such that the Cas9-derived protein lacks all nuclease activity. Also, either the RuvC-like domain or the HNH-like domain can be inactivated independently of each other to yield a protein with nickase activity.

[0138] In any of the above-described embodiments, any or all of the nuclease domains can be inactivated by one or more deletion mutations, insertion mutations, and/or substitution mutations using

well-known methods, such as site-directed mutagenesis, PCR-mediated mutagenesis, and total gene synthesis, as well as other meths known in the art.

[0139] **Nucleic Acids Encoding RNA-Guided Endonucleases or Fusion Proteins**

[0140] Another aspect of the present disclosure provides nucleic acids encoding any of the RNA-guided endonucleases or fusion proteins described above in sections (I) and (II), respectively. The nucleic acid can be RNA or DNA. In one embodiment, the nucleic acid encoding the RNA-guided endonuclease or fusion protein is mRNA. The mRNA can be 5' capped and/or 3' polyadenylated. In another embodiment, the nucleic acid encoding the RNA-guided endonuclease or fusion protein is DNA. The DNA can be present in a vector (see below).

[0141] The nucleic acid encoding the RNA-guided endonuclease or fusion protein can be codon optimized for efficient translation into protein in the cell, plant or animal of interest. For example, codons can be optimized for expression in humans, mice, rats, hamsters, cows, pigs, cats, dogs, fish, amphibians, plants, yeast, insects, and so forth. Programs for codon optimization are available as freeware. Commercial codon optimization programs are also available.

[0142] In some embodiments, DNA encoding the RNA-guided endonuclease or fusion protein can be operably linked to at least one promoter control sequence. In some iterations, the DNA coding sequence can be operably linked to a promoter control sequence for expression in the eukaryotic cell or animal of interest. The promoter control sequence can be constitutive, regulated, or tissue-specific. Suitable constitutive promoter control sequences include, but are not limited to, cytomegalovirus immediate early promoter (CMV), simian virus (SV40) promoter, adenovirus major late promoter, Rous sarcoma virus (RSV) promoter, mouse mammary tumor virus (MMTV) promoter, phosphoglycerate kinase (PGK) promoter, elongation factor (EF1)-alpha promoter, ubiquitin promoters, actin promoters, tubulin promoters, immunoglobulin promoters, fragments thereof, or

combinations of any of the foregoing. Examples of suitable regulated promoter control sequences include without limit those regulated by heat shock, metals, steroids, antibiotics, or alcohol. Non-limiting examples of tissue-specific promoters include B29 promoter, CD14 promoter, CD43 promoter, CD45 promoter, CD68 promoter, desmin promoter, elastase-1 promoter, endoglin promoter, fibronectin promoter, Flt-1 promoter, GFAP promoter, GPIIb promoter, ICAM-2 promoter, INF-β promoter, Mb promoter, Nphsl promoter, OG-2 promoter, SP-B promoter, SYN1 promoter, and WASP promoter. The promoter sequence can be wild type or it can be modified for more efficient or efficacious expression. In one exemplary embodiment, the encoding DNA can be operably linked to a CMV promoter for constitutive expression in mammalian cells.

[0143] In certain embodiments, the sequence encoding the RNA-guided endonuclease or fusion protein can be operably linked to a promoter sequence that is recognized by a phage RNA polymerase for *in vitro* mRNA synthesis. In such embodiments, the *in vitro*-transcribed RNA can be purified for use in the methods detailed below in sections (IV) and (V). For example, the promoter sequence can be a T7, T3, or SP6 promoter sequence or a variation of a T7, T3, or SP6 promoter sequence. In an exemplary embodiment, the DNA encoding the fusion protein is operably linked to a T7 promoter for in vitro mRNA synthesis using T7 RNA polymerase.

[0144] In alternate embodiments, the sequence encoding the RNA-guided endonuclease or fusion protein can be operably linked to a promoter sequence for *in vivo* expression of the RNA-guided endonuclease or fusion protein in bacterial or eukaryotic cells. In such embodiments, the expressed protein can be purified for use in the methods detailed below in sections (IV) and (V). Suitable bacterial promoters include, without limit, T7 promoters, lac operon promoters, trp promoters, variations thereof, and combinations thereof. An exemplary bacterial promoter is tac which is a hybrid of trp and lac promoters. Non-limiting examples of suitable eukaryotic promoters are listed above.

[0145] In additional aspects, the DNA encoding the RNA-guided endonuclease or fusion protein also can be linked to a polyadenylation signal (*e.g.*, SV40 polyA signal, bovine growth hormone (BGH) polyA signal, etc.) and/or at least one transcriptional termination sequence. Additionally, the sequence encoding the RNA-guided endonuclease or fusion protein also can be linked to sequence encoding at least one nuclear localization signal, at least one cell-penetrating domain, and/or at least one marker domain, which are detailed above in section (I).

[0146] In various embodiments, the DNA encoding the RNA-guided endonuclease or fusion protein can be present in a vector. Suitable vectors include plasmid vectors, phagemids, cosmids, artificial/mini-chromosomes, transposons, and viral vectors (*e.g.*, lentiviral vectors, adeno-associated viral vectors, etc.). In one embodiment, the DNA encoding the RNA-guided endonuclease or fusion protein is present in a plasmid vector. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. The vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, antibiotic resistance genes), origins of replication, and the like. Additional information can be found in "Current Protocols in Molecular Biology" Ausubel, *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3rd edition, 2001.

[0147] In some embodiments, the expression vector comprising the sequence encoding the RNA-guided endonuclease or fusion protein can further comprise sequence encoding a gRNA. The sequence encoding the gRNA generally is operably linked to at least one transcriptional control sequence for expression of the gRNA in the cell or embryo of interest. For example, DNA encoding the gRNA can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6, U3, H1, and 7SL RNA promoters.

[0148] **Methods for Modifying a Chromosomal Sequence Using an RNA-Guided Endonuclease or Site-Specific Endonuclease**

[0149] Another aspect of the present disclosure encompasses a method for modifying a chromosomal sequence in a eukaryotic cell or embryo. The method for base editing comprises introducing into a eukaryotic cell or embryo (i) at least one base editor comprising (a) a partially or fully enzymatically inactivated RNA-guided endonuclease, (b) at least one catalytic effector domain, and, optionally, (c) at least one nuclear localization signal, or nucleic acid encoding said base editor, (ii) at least one gRNA or DNA encoding at least one gRNA, and, optionally, (iii) a uracil DNA glycosylase inhibitor (UGI). The catalytic effector domain may be a cytosine deaminase, an adenine deaminase, or a guanine-specific glycosylase. The method for base editing further comprises culturing the cell or embryo such that each gRNA directs an partially or fully enzymatically inactivated RNA-guided endonuclease to a targeted site in the chromosomal sequence where the catalytic effector domain modifies or excises a nucleobase at the targeted site, and the modified or excised nucleobase is repaired (*i.e.*, converted) by a DNA repair process such that the chromosomal sequence is modified. The method for homology directed repair comprises introducing into a eukaryotic cell or embryo (i) at least one RNA-guided endonuclease (or encoding nucleic acid) comprising at least one nuclear localization signal or nucleic acid encoding at least one RNA-guided endonuclease comprising at least one nuclear localization signal, (ii) at least one gRNA or DNA encoding at least one gRNA, and, (iii) at least one DNA donor sequence. The donor sequence can be a single or double stranded DNA oligonucleotide or plasmid, with regions of homology to the sequence targeted by the gRNA. The method for homology directed repair further comprises culturing the cell or embryo such that each gRNA directs an RNA-guided endonulease to a targeted site in the chromosomal sequence where the endonuclease cleaves or nicks at the targeted site, and the double or single stranded

break is repaired by a DNA repair process using the provided DNA donor sequence such that the chromosomal sequence is modified.

[0150] Accordingly, as discussed herein, the targeted chromosomal sequence can be modified or inactivated. For example, using base editing or HDR a single nucleotide change (SNP) can give rise to an altered protein product or introduce a “stop” codon into the reading frame of a coding sequence and thereby inactivate or "knock out" the sequence such that no protein product is made. As another example, using HDR a defined sequence can be inserted in (or "knocked in") to a gene sequence. This defined sequence may include functional DNA sequences, including but not limited to transcription factor binding sites, constitutive or inducible promoters, reporter genes, and/or new exons.

[0151] In other embodiments, the method can comprise introducing two (or more) base editors or RNA-guided endonucleases (or encoding nucleic acid) and two or more gRNAs (or encoding DNA) into a cell or embryo, wherein the base editors modify nucleobases at two or more sites or the RNA-guided endonucleases create double or single stranded breaks at two or more sites. The modified nucleobases or DNA breaks can be within several base pairs, within tens of base pairs, or can be separated by many thousands of base pairs.

[0152] **(a) RNA-Guided Endonuclease**

[0153] The methods comprise introducing into a cell or embryo at least one RNA-guided endonuclease comprising at least one nuclear localization signal or nucleic acid encoding at least one RNA-guided endonuclease comprising at least one nuclear localization signal. Such RNA-guided endonucleases and nucleic acids encoding RNA-guided endonucleases are described above in sections (I) and (III), respectively. Such RNA-guided endonucleases may be preloaded with a gRNA.

[0154] In some embodiments, the RNA-guided endonuclease can be introduced into the cell or embryo as an isolated protein. In such embodiments, the RNA-guided endonuclease can further comprise at least one cell-penetrating domain, which facilitates cellular uptake of the protein. In other embodiments, the RNA-guided endonuclease can

be introduced into the cell or embryo as an mRNA molecule. In still other embodiments, the RNA-guided endonuclease can be introduced into the cell or embryo as a DNA molecule. In general, DNA sequence encoding the fusion protein is operably linked to a promoter sequence that will function in the cell or embryo of interest. The DNA sequence can be linear, or the DNA sequence can be part of a vector. In still other embodiments, the fusion protein can be introduced into the cell or embryo as an RNA-protein complex comprising the fusion protein and the gRNA.

[0155] In alternate embodiments, DNA encoding the RNA-guided endonuclease can further comprise sequence encoding a gRNA. In general, each of the sequences encoding the RNA-guided endonuclease and the gRNA is operably linked to an appropriate promoter control sequence that allows expression of the RNA-guided endonuclease and the gRNA, respectively, in the cell or embryo. The DNA sequence encoding the RNA-guided endonuclease and the gRNA can further comprise additional expression control, regulatory, and/or processing sequence(s). The DNA sequence encoding the RNA-guided endonuclease and the gRNA can be linear or can be part of a vector

[0156] **(b) Guide RNA (gRNA)**

[0157] The methods also comprise introducing into a cell or embryo at least one gRNA or DNA encoding at least one gRNA. A gRNA interacts with theRNA-guided endonuclease to direct the endonuclease to a specific target site in the chromosomal sequence.

[0158] Each gRNA may comprise three regions: a first region that is complementary to the target site in the chromosomal sequence, a second internal region that forms a stem loop structure, and a third region that remains essentially single-stranded. The first region of each gRNA is different such that each gRNA guides a fusion protein to a specific target site. The second and third regions of each gRNA can be the same in all gRNAs.

[0159] The first region of the gRNA, termed the spacer, is complementary to sequence at the target site in the chromosomal

sequence (*i.e.*, protospacer sequence) such that the first region of the gRNA can base pair with the target site. In various embodiments, the first region of the gRNA can comprise from about 10 nucleotides to more than about 25 nucleotides. For example, the region of base pairing between the first region of the gRNA and the target site in the chromosomal sequence can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more than 25 nucleotides in length. In an exemplary embodiment, the first region of the gRNA is about 19, 20, 21, 22 or 23 nucleotides in length.

[0160] The gRNA also comprises a second region that forms a secondary structure. In some embodiments, the secondary structure comprises a stem (or hairpin) and a loop. The length of the loop and the stem can vary. For example, the loop can range from about 3 to about 10 nucleotides in length, and the stem can range from about 6 to about 20 base pairs in length. The stem can comprise one or more bulges of 1 to about 10 nucleotides. Thus, the overall length of the second region can range from about 16 to about 60 nucleotides in length. In an exemplary embodiment, the loop is about 4 nucleotides in length and the stem comprises about 12 base pairs.

[0161] The gRNA may also comprise a third region that remains essentially single-stranded. Thus, the third region has no complementarity to any chromosomal sequence in the cell of interest and has no complementarity to the rest of the gRNA. The length of the third region can vary. In general, the third region is more than about 4 nucleotides in length. For example, the length of the third region can range from about 5 to about 60 nucleotides in length.

[0162] The combined length of the second and third regions (also called the universal or scaffold region) of the gRNA can range from about 20 to about 120 nucleotides in length. In one aspect, the combined length of the second and third regions of the gRNA range from about 20 to about 100 nucleotides in length.

[0163] In some embodiments, the gRNA comprises a single molecule comprising all three regions, known as an sgRNA. In other embodiments, the gRNA can comprise two separate molecules. The

first RNA molecule can comprise the first region of the gRNA and one half of the "stem" of the second region of the gRNA, known as the crRNA. The second RNA molecule can comprise the other half of the "stem" of the second region of the gRNA and the third region of the gRNA, known as the tracrRNA. Thus, in this embodiment, the first and second RNA molecules each contain a sequence of nucleotides that are complementary to one another. For example, in one embodiment, the first and second RNA molecules each comprise a sequence (of about 6 to about 20 nucleotides) that base pairs to the other sequence to form a functional gRNA.

[0164] In some embodiments, the gRNA can be introduced into the cell or embryo as an RNA molecule. The RNA molecule can be transcribed *in vitro*. Alternatively, the RNA molecule can be chemically synthesized.

[0165] In other embodiments, the gRNA can be introduced into the cell or embryo as a DNA molecule. In such cases, the DNA encoding the gRNA can be operably linked to promoter control sequences for expression of the gRNA in the cell or embryo of interest. For example, the RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6 or H1 promoters. In exemplary embodiments, the RNA coding sequence is linked to a mouse or human U6 promoter. In other exemplary embodiments, the RNA coding sequence is linked to a mouse or human H1 promoter.

[0166] The DNA molecule encoding the gRNA can be linear or circular. In some embodiments, the DNA sequence encoding the gRNA can be part of a vector. Suitable vectors include plasmid vectors, phagemids, cosmids, artificial/mini-chromosomes, transposons, and viral vectors. In an exemplary embodiment, the DNA encoding the gRNA is present in a plasmid vector. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. The vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences,

transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, antibiotic resistance genes), origins of replication, and the like.

[0167] In embodiments in which both the RNA-guided endonuclease and the gRNA are introduced into the cell as DNA molecules, each can be part of a separate molecule (*e.g.*, one vector containing fusion protein coding sequence and a second vector containing gRNA coding sequence) or both can be part of the same molecule (*e.g.*, one vector containing coding and regulatory sequence for both the fusion protein and the gRNA).

[0168] **(c) Target Site**

[0169] An RNA-guided endonuclease in conjunction with a gRNA is directed to a target site in the chromosomal sequence, wherein the RNA-guided endonuclease binds the chromosomal sequence. The target site has no sequence limitation except that the sequence is immediately adjacent to a consensus sequence. This consensus sequence is also known as a protospacer adjacent motif (PAM). Examples of PAM include, but are not limited to, NGG (SEQ ID NO: 18), NGGNG (SEQ ID NO: 19), and NNAGAAW (SEQ ID NO: 20) (wherein N is defined as any nucleotide and W is defined as either A or T). As detailed above in section (IV)(b), the spacer region of the gRNA is complementary to the protospacer of the target sequence. Typically, the spacer region of the gRNA is about 19 to 21 nucleotides in length. Thus, in certain aspects, the sequence of the target site in the chromosomal sequence is 5'-N_{19-21}-NGG-3'.

[0170] The target site can be in the coding region of a gene, in an intron of a gene, in a control region of a gene, in a non-coding region between genes, etc. The gene can be a protein coding gene or an RNA coding gene. The gene can be any gene of interest.

[0171] **(d) Introducing Into the Cell or Embryo**

[0172] The RNA-targeted endonuclease(s) (or encoding nucleic acid) and the gRNA(s) (or encoding DNA) can be introduced into a cell or embryo by a variety of means. In some embodiments, the cell or embryo is transfected. Suitable transfection methods include calcium

phosphate-mediated transfection, nucleofection (or electroporation), cationic polymer transfection (*e.g.*, DEAE-dextran or polyethylenimine), viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, gene gun delivery, impalefection, sonoporation, optical transfection, and proprietary agent-enhanced uptake of nucleic acids. Transfection methods are well known in the art (see, *e.g.*, "Current Protocols in Molecular Biology" Ausubel, *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3rd edition, 2001). In other embodiments, the molecules are introduced into the cell or embryo by microinjection. Typically, the embryo is a fertilized one-cell stage embryo of the species of interest. For example, the molecules can be injected into the pronuclei of one cell embryos.

[0173] The RNA-guided endonuclease(s) (or encoding nucleic acid), the gRNA(s) (or DNAs encoding the gRNA), and the optional uracil glycosylase inhibitor can be introduced into the cell or embryo simultaneously or sequentially. The ratio of the RNA-guided endonuclease(s) (or encoding nucleic acid) to the gRNA(s) (or encoding DNA) generally will be about stoichiometric such that they can form an RNA-protein complex. In one embodiment, DNA encoding an RNA-targeted endonuclease and DNA encoding a gRNA are delivered together within the plasmid vector.

[0174] **(e) Culturing the Cell or Embryo**

[0175] The method further comprises maintaining the cell or embryo under appropriate conditions such that the gRNA(s) directs the base editor fusion protein(s) to the targeted site(s) in the chromosomal sequence, and the catalytic effector domain introduces at least one nucleobase conversion in the chromosomal sequence such that the chromosomal sequence is modified by a substitution of at least one nucleotide.

[0176] In embodiments in which a premature stop codon is introduced, the chromosomal sequence may be inactivated or "knocked out." An inactivated protein-coding chromosomal sequence does not give rise to the protein coded by the wildtype chromosomal sequence.

[0177] In embodiments in which the chromosome sequence is modified without knocking out or inactivating the sequence, the protein encoded may be either 1) modified resulting in a protein with lesser or greater function or 2) in the case of correcting natural mutations, brought back to or approximate wildtype sequence and/or function. In general, the cell is maintained under conditions appropriate for cell growth and/or maintenance. Suitable cell culture conditions are well known in the art and are described, for example, in Santiago, *et al.*, (2008) PNAS 105:5809-5814; Moehle, *et al.*, (2007) PNAS 104:3055-3060; Urnov, *et al.*, (2005) Nature 435:646-651; and Lombardo, *et al.* (2007) Nat. Biotechnology 25:1298-1306. Those of skill in the art appreciate that methods for culturing cells are known in the art and can and will vary depending on the cell type. Routine optimization may be used, in all cases, to determine the best techniques for a particular cell type.

[0178] An embryo can be cultured *in vitro* (*e.g.*, in cell culture). Typically, the embryo is cultured at an appropriate temperature and in appropriate media with the necessary O_2/CO_2 ratio to allow for editing to take place. Suitable non-limiting examples of media include M2, M16, KSOM, BMOC, and HTF media. A skilled artisan will appreciate that culture conditions can and will vary depending on the species of embryo. Routine optimization may be used, in all cases, to determine the best culture conditions for a particular species of embryo. In some cases, a cell line may be derived from an *in vitro*-cultured embryo (*e.g.*, an embryonic stem cell line).

[0179] Alternatively, an embryo may be cultured *in vivo* by transferring the embryo into the uterus of a female host. Generally speaking, the female host is from the same or similar species as the embryo. Preferably, the female host is pseudo-pregnant. Methods of preparing pseudo-pregnant female hosts are known in the art. Additionally, methods of transferring an embryo into a female host are known.

Culturing an embryo *in vivo* permits the embryo to develop and can result in a live birth of an animal derived from the embryo. Such an animal would comprise the modified chromosomal sequence in every cell of the body.

[0180] **(f) Cell and Embryo Types**

[0181] A variety of eukaryotic cells and embryos are suitable for use in the method. For example, the cell can be a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, an insect cell, a plant cell, a yeast cell, a single cell eukaryotic organism, or a prokaryote. In general, the embryo is a non-human mammalian embryo. In specific embodiments, the embryos can be a one cell non-human mammalian embryo. Exemplary mammalian embryos, including one cell embryos, include without limit mouse, rat, hamster, rodent, rabbit, feline, canine, ovine, porcine, bovine, equine, and nonhuman primate embryos. In still other embodiments, the cell can be a stem cell. Suitable stem cells include without limit embryonic stem cells, ES-like stem cells, fetal stem cells, adult stem cells, pluripotent stem cells, induced pluripotent stem cells, multipotent stem cells, oligopotent stem cells, unipotent stem cells and others. In exemplary embodiments, the cell is a mammalian cell.

[0182] Non-limiting examples of suitable mammalian cells include Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells; mouse myeloma NSO cells, mouse embryonic fibroblast 3T3 cells (NIH3T3), mouse B lymphoma A20 cells; mouse melanoma B16 cells; mouse myoblast C2C12 cells; mouse myeloma SP2/0 cells; mouse embryonic mesenchymal C3H-10T1/2 cells; mouse carcinoma CT26 cells, mouse prostate DuCuP cells; mouse breast EMT6 cells; mouse hepatoma Hepa1c1c7 cells; mouse myeloma J5582 cells; mouse epithelial MTD-1A cells; mouse myocardial MyEnd cells; mouse renal RenCa cells; mouse pancreatic RIN-5F cells; mouse melanoma X64 cells; mouse lymphoma YAC-1 cells; rat glioblastoma 9L cells; rat B lymphoma RBL cells; rat neuroblastoma B35 cells; rat hepatoma cells (HTC); buffalo rat liver BRL 3A cells; canine kidney cells (MDCK); canine mammary (CMT) cells; rat osteosarcoma D17 cells; rat

monocyte/macrophage DH82 cells; monkey kidney SV-40 transformed fibroblast (COS7) cells; monkey kidney CVI-76 cells; African green monkey kidney (VERO-76) cells; human embryonic kidney cells (HEK293, HEK293T); human cervical carcinoma cells (HELA); human lung cells (W138); human liver cells (Hep G2); human U2-OS osteosarcoma cells, human A549 cells, human A-431 cells, and human K562 cells. An extensive list of mammalian cell lines may be found in the American Type Culture Collection catalog (ATCC, Manassas, Va.).

[0183] **(g) Donor Sequence**

[0184] The HDR method also comprises introducing into a cell or embryo at least one DNA donor sequence. The donor sequence is used by the cell's DNA repair mechanisms as a template for repair at the site that was nicked or cleaved by the endonuclease at the specific target site encoded by the gRNA sequence.

[0185] Each DNA donor may be single or double stranded and may be a linear or circular oligonucleotide. Each donor comprises two homologus regions, termed homology arms, that are complementary to the regions flanking the target site in the chromosomal sequence. Each donor may also comprise the specific sequence change or sequence insertion, located between the two homology arms.

[0186] The homology arms are complementary to the sequence flanking the gRNA target site in the chromosomal sequence, such that the homology arms are recognized by the cell as a single stranded or double stranded copy of the chromosomal sequence that was cleaved by the endonuclease. In various embodiments, the homology arms can comprise from about 25 nucleotides to over 1000 nucleotides, depending on the size of the specific sequence change or insertion desired. For example, for single base pair changes or insertions the homology arms can be from 25-50 nucleotides in length, but may be longer. Alternatively, for larger sequence insertions, including but not limited to fluorescent proteins, the homology arms can be between 500-1000 nucleotides in length, but may be longer or shorter with varying effectiveness. One of skill in the art can determine the optimal length of the homology arms for a particular use.

[0187] In some embodiments, the donor is a single stranded linear DNA oligonucleotide, known as an ssODN. In other embodiments, the donor is a double stranded linear DNA oligonucleotide, known as a dsODN. In some embodiments these ODNs are artificially synthesized and can be modified chemically or by sequence for stability, on target capabilities, and/or integration efficiency. In other embodiments they are produced by PCR. In an exemplary embodiment, the DNA donor is a synthesized ssODN.

[0188] In some embodiments, the donor is a single or double stranded circular DNA oligonucleotide. In some embodiments, the DNA sequence can be part of a vector. Suitable vectors include plasmid vectors, artificial/mini-chromosomes, and viral vectors. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. The vector can comprise additional expression control sequences (e.g., enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, antibiotic resistance genes), origins of replication, and the like. Non-limiting examples of viral vectors include lentivirus and adeno-associated virus (AAV).

[0189] In embodiments in which the RNA-guided endonuclease, the gRNA, and the donor are introduced into the cell as DNA molecules, each can be introduced as a separate molecule (*e.g.*, one vector containing protein coding sequence, a second vector containing gRNA coding sequence, and a third vector containing the DNA donor sequence), or all three can be part of the same molecule (*e.g.*, one vector containing coding and regulatory sequence for the protein, the gRNA, and the donor template). Alternatively, any two can be part of the same DNA molecule. For example, both the protein coding sequence and gRNA coding sequence can be part of the same molecule, and the DNA donor sequence can be a second molecule, or any combination of vectors. (*e.g.*, one vector containing coding and regulatory sequence for both the fusion protein and the gRNA, and a second vector or ODN containing the DNA donor sequence).

[0190] **(V) Method for Using a Fusion Protein to Modify a Chromosomal Sequence or Regulate Expression of a Chromosomal Sequence**

[0191] Another aspect of the present disclosure encompasses a method for modifying a chromosomal sequence or regulating expression of a chromosomal sequence in a cell or embryo. The method comprises introducing into the cell or embryo (a) at least one base editor fusion protein or nucleic acid encoding at least one fusion protein, wherein the fusion protein comprises a CRISPR/Cas-like protein or a fragment thereof and a catalytic effector domain or tethered protein encoding a catalytic effector domain, and (b) at least one gRNA or DNA encoding the gRNA, wherein the gRNA guides the CRISPR/Cas-like protein of the fusion protein to a targeted site in the chromosomal sequence and the catalytic effector domain of the fusion protein modifies the chromosomal sequence or regulates expression of the chromosomal sequence by modification of a regulatory sequence associated with the chromosomal sequence.

[0192] Fusion proteins comprising a CRISPR/Cas-like protein or a fragment thereof and a catalytic effector domain are detailed above in section (II). In general, if used with fusion proteins, the fusion proteins may comprise at least one nuclear localization signal. Nucleic acids encoding fusion proteins are described above in section (III). In some embodiments, the fusion protein can be introduced into the cell or embryo as an isolated protein (which can further comprise a cell-penetrating domain). Furthermore, the isolated fusion protein can be part of a protein-RNA complex comprising the gRNA. In other embodiments, the fusion protein can be introduced into the cell or embryo as an RNA molecule (which can be capped and/or polyadenylated). In still other embodiments, the fusion protein can be introduced into the cell or embryo as a DNA molecule. For example, the fusion protein and the gRNA can be introduced into the cell or embryo as discrete DNA molecules or as part of the same DNA molecule. Such DNA molecules can be plasmid vectors.

[0193] **(VI) Genetically Modified Cells and Animals**

[0194] The present disclosure encompasses genetically modified cells, non-human embryos, and non-human animals comprising at least one chromosomal sequence that has been modified using an RNA-guided endonuclease-mediated or fusion protein-mediated process, for example, using the methods described herein. The disclosure provides cells comprising: (i) at least one DNA or RNA molecule encoding an RNA-guided endonuclease or fusion protein targeted to a chromosomal sequence of interest or a fusion protein, (ii) at least one gRNA, for HDR (iii) a DNA donor sequence, and, optionally for base editing, (iv) one or more free standing or fused uracil glycosylase inhibitors (UGI). The disclosure also provides non-human embryos comprising: (i) at least one DNA or RNA molecule encoding an enzymatically inactivated RNA-guided endonuclease or fusion protein targeted to a chromosomal sequence of interest, (ii) at least one gRNA, for HDR (iii) a DNA donor sequence, and, optionally for base editing, (iv) one or more UGIs.

[0195] The present disclosure provides genetically modified non-human animals, non-human embryos, or animal cells comprising at least one modified chromosomal sequence. The modified chromosomal sequence may be modified such that it is (1) inactivated or (2) has an altered expression or produces an altered protein product. The chromosomal sequence is modified with an RNA guided endonuclease-mediated or fusion protein-mediated process, using the methods described herein.

[0196] As discussed, one aspect of the present disclosure provides a genetically modified animal in which at least one chromosomal sequence has been modified. In one embodiment, the genetically modified animal comprises at least one inactivated chromosomal sequence. The modified chromosomal sequence may be inactivated such that the sequence is not transcribed and/or a functional protein product is not produced. Thus, a genetically modified animal comprising an inactivated chromosomal sequence may be termed a "knock out." As a consequence of the mutation, the targeted chromosomal sequence is inactivated and a functional protein is not

produced. The inactivated chromosomal sequence comprises no exogenously introduced sequence. Also included herein are genetically modified animals in which two, three, four, five, six, seven, eight, nine, or ten or more chromosomal sequences are inactivated.

[0197] In another embodiment, the modified chromosomal sequence can be altered such that it codes for a variant protein product. For example, a genetically modified animal comprising a modified chromosomal sequence can comprise a targeted point mutation(s) or other modification such that an altered protein product is produced. In one embodiment, the chromosomal sequence can be modified such that at least one nucleotide is changed, and the expressed protein comprises one changed amino acid residue (missense mutation). In another embodiment, the chromosomal sequence can be modified to comprise more than one missense mutation such that more than one amino acid is changed. The altered or variant protein can have altered properties or activities compared to the wildtype protein (or as compared to an unmodified protein in the case of correcting a natural genetic mutation), such as altered substrate specificity, altered enzyme activity, altered kinetic rates, etc.

[0198] In another embodiment, the genetically modified animal can comprise at least one genetic change where a non-expressed gene may be activated or inserted, which is termed a "knock in." The chromosomally modified sequence can, for example, encode a protein resembling an orthologous protein, an endogenous protein, or combinations of both.

[0199] In yet another embodiment, the genetically modified animal can comprise at least one modified chromosomal sequence encoding a protein such that the expression pattern of the protein is altered. For example, regulatory regions controlling the expression of the protein, such as a promoter or a transcription factor binding site, can be altered such that the protein is over-produced, or the tissue-specific or temporal expression of the protein is altered, or a combination thereof. Alternatively, the expression pattern of the protein can be altered in combination with using a conditional knockout system. A non-limiting

example of a conditional knockout system includes a Cre-lox recombination system. A Cre-lox recombination system comprises a Cre recombinase enzyme, a site-specific DNA recombinase that can catalyze the recombination of a nucleic acid sequence between specific sites (lox sites) in a nucleic acid molecule. Methods of using this system to produce temporal and tissue specific expression are known in the art. In general, a genetically modified animal is generated with lox sites flanking a chromosomal sequence. The genetically modified animal comprising the lox-flanked chromosomal sequence can then be crossed with another genetically modified animal expressing Cre recombinase. Progeny animals comprising the lox-flanked chromosomal sequence and the Cre recombinase are then produced, and the lox-flanked chromosomal sequence is recombined, leading to deletion or inversion of the chromosomal sequence encoding the protein. Expression of Cre recombinase can be temporally and conditionally regulated to effect temporally and conditionally regulated recombination of the chromosomal sequence.

[0200] In any of these embodiments, the genetically modified animal disclosed herein can be heterozygous for the modified chromosomal sequence. Alternatively, the genetically modified animal can be homozygous for the modified chromosomal sequence.

[0201] The genetically modified animals disclosed herein can be crossbred to create animals comprising more than one modified chromosomal sequence or to create animals that are homozygous for one or more modified chromosomal sequences. For example, two animals comprising the same modified chromosomal sequence can be crossbred to create an animal homozygous for the modified chromosomal sequence. Alternatively, animals with different modified chromosomal sequences can be crossbred to create an animal comprising both modified chromosomal sequences.

[0202] In other embodiments, an animal comprising a modified chromosomal sequence can be crossbred to combine the modified chromosomal sequence with other genetic backgrounds. By way of non-limiting example, other genetic backgrounds may include wild-type

genetic backgrounds, genetic backgrounds with deletion mutations, genetic backgrounds with a targeted integration, and genetic backgrounds with non-targeted integrations.

[0203] The term "animal," as used herein, refers to a human or non-human animal. The animal may be an embryo, a juvenile, or an adult. Suitable animals include vertebrates such as mammals, birds, reptiles, amphibians, and fish. Examples of suitable mammals include without limit rodents, companion animals, livestock, and primates. Non-limiting examples of rodents include mice, rats, hamsters, gerbils, and guinea pigs. Suitable companion animals include but are not limited to cats, dogs, rabbits, hedgehogs, and ferrets. Non-limiting examples of livestock include horses, goats, sheep, swine, cattle, llamas, and alpacas. Suitable primates include but are not limited to capuchin monkeys, chimpanzees, lemurs, macaques, marmosets, tamarins, spider monkeys, squirrel monkeys, and vervet monkeys. Non-limiting examples of birds include chickens, turkeys, ducks, and geese. Alternatively, the animal may be an invertebrate such as insects, nematodes, shellfish, and the like. Non-limiting examples of insects include Drosophila and mosquitoes. An exemplary animal is a rat. Non-limiting examples of suitable rat strains include Dahl Salt-Sensitive, Fischer 344, Lewis, Long Evans Hooded, Sprague-Dawley, and Wistar. In one embodiment, the animal is not a genetically modified mouse. In each of the foregoing iterations of suitable animals for the invention, the animal does not include exogenously introduced, randomly integrated transposon sequences.

[0204] A further aspect of the present disclosure provides genetically modified cells or cell lines comprising at least one modified chromosomal sequence. The genetically modified cell or cell line can be derived from any of the genetically modified animals disclosed herein. Alternatively, the chromosomal sequence can be modified in a cell as described herein above (in the paragraphs describing chromosomal sequence modifications in animals) using the methods described herein. The disclosure also encompasses a lysate of said cells or cell lines.

[0205] In preferred embodiments, the cells are eukaryotic cells. Suitable host cells include fungi or yeast, such as *Pichia*, *Saccharomyces*, or *Schizosaccharomyces*; insect cells, such as SF9 cells from *Spodoptera frugiperda* or S2 cells from *Drosophila melanogaster*; and animal cells, such as mouse, rat, hamster, non-human primate, or human cells. Exemplary cells are mammalian. The mammalian cells can be primary cells. The cells may be of a variety of cell types, *e.g.*, fibroblast, myoblast, T or B cell, macrophage, epithelial cell, and so forth.

[0206] The cells can be eukaryotic cells or prokaryotic cells. For example, cells from bacteria, protists, plants, animals, fungi are contemplated. Animals may include, but are not limited to, insects and mammals.

[0207] When mammalian cell lines are used, the cell line can be any established cell line or primary cell type, or one that is not yet described. The cell line can be adherent or non-adherent, or the cell line can be grown under conditions that encourage adherent, non-adherent or organotypic growth using standard techniques known to individuals skilled in the art. Non-limiting examples of suitable mammalian cells and cell lines are provided herein in section (IV)(g). In still other embodiments, the cell can be a stem cell. Non-limiting examples of suitable stem cells are provided in section (IV)(g).

[0208] The present disclosure also provides a genetically modified non-human embryo comprising at least one modified chromosomal sequence. The chromosomal sequence can be modified in an embryo as described herein above (in the paragraphs describing chromosomal sequence modifications in animals) using the methods described herein. In one embodiment, the embryo is a non-human fertilized one-cell stage embryo of the animal species of interest. Exemplary mammalian embryos, including one cell embryos, include without limit, mouse, rat, hamster, rodent, rabbit, feline, canine, ovine, porcine, bovine, equine, and primate embryos.

[0209] Further, non-mammalian cells and cell lines may be used including, but not limited to, plant, insect and prokaryote cells and cell

lines, as are known to one of skill in the art. Such cells may be newly developed for a specific purpose or may be readily available such as from ATCC (Bethesda, MD) and other sources known to one of skill in the art. In short, the compositions and methods of the present invention should be useful for base editing in any cell having nucleic acids and Cas-based repair pathways including viruses after infection into a host cell.

[0210] **(VII) Kits**

[0211] Still another aspect of the present disclosure provides kits for carrying out the methods described above.

[0212] The kits provided herein generally include instructions for carrying out the processes detailed above. Instructions included in the kits may be affixed to packaging material, may be included as a package insert or as a downloadable file. While the instructions are typically written or printed materials, they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this disclosure. Such media include, but are not limited to, electronic storage media (*e.g.*, magnetic discs, tapes, cartridges, chips), optical media (*e.g.*, CD ROM), and the like. As used herein, the term "instructions" can include the address of an internet site that provides the instructions.

[0213] As various changes could be made in the above-described processes and kits without departing from the scope of the invention, it is intended that all matter contained in the above description and in the examples given below, shall be interpreted as illustrative and not in a limiting sense.

Exemplification

[0214] **Example 1**

[0215] **Simultaneous inhibition of SSBR and NHEJ increase the rate of C-to-T editing and decrease the rate of indel formation**

[0216] The PARP1/2 inhibitor AZD2281 (olaparib) and the DNA dependent protein kinase (DNA-PK) inhibitor MSC2528072A were

obtained from Merck Healthcare (Darmstadt, DE) as 10 mM solutions in DMSO.

[0217] Cytosine base editor (CBE) protein was constructed by fusing the C-terminal domain of human APOBEC3B to the amino-terminus of an SpCas9 nickase protein (SEQ ID NO: 1). The protein was expressed and purified from *E. coli* BL21AI by autoinduction, nickel column chromatography, and cation exchange chromatography and was stored at –80℃ before use in a buffer containing 10% glycerol, 300 mM KCl, 20 mM HEPES (pH 7.5), and 1 mM DTT. Synthetic single guide RNAs (sgRNAs) were purchased from MilliporeSigma (Burlington, MA). The spacer sequences of the sgRNA are listed in Table 1. Each experimental condition was tested in two technical replicates.

Table 1. sgRNA spacer sequences and PCR primer sequences

	Sequence (5'-3')
EMX1-11 spacer	GGAGCCCTTCTTCTTCTGCT (SEQ ID NO: 3)
HBB spacer	CTTGCCCCACAGGGCAGTAA (SEQ ID NO: 4)
EMX1-11 PCR forward	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNAAGCAGCACTCTGCCCTCGT (SEQ ID NO: 5)
EMX1-11 PCR reverse	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNCCCTATGTAGCCTCAGTCTTCC (SEQ ID NO: 6)
HBB PCR forward	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNGCCAATCTACTCCCAGGAGC (SEQ ID NO: 7)
HBB PCR reverse	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNAGGCAGAGAGAGTCAGTGCCTA (SEQ ID NO: 8)

[0218] Ribonucleoprotein (RNP) complexes were prepared by incubating for 15 minutes at room temperature: 76 pmol of CBE protein

and 230 pmol of sgRNA in buffer (20 mM HEPES, 100 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, pH 7.5) for a final volume of 10 µL. RNPs were kept on ice until transfection. HEK293 cells were obtained from ATCC (Bethesda, MD) and grown at 37 ℃ and 5% CO_2 in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 1 mM sodium pyruvate, and 0.1 mM non-essential amino acids. Cells were seeded at 1.67 x 10^4 cells/cm^2 of tissue culture surface area two days before transfection. At the time of transfection, cells were trypsinized to obtain a single-cell suspension, washed twice with Hank's Balanced Salt Solution and resuspended in Nucleofector Solution V (Lonza, Basel, CH) at 10^6 cells per 100 µL. Nucleofection was performed by mixing 100 µL of prepared cell suspension with 10 µL of complexed CBE RNP by pipetting up and down six times before transferring to a cuvette for electroporation using program Q-001 on a Nucleofector 2b machine (Lonza, Basel, CH). Nucleofected cells were immediately transferred to 12-well plates at a density of 10^5 cells per well in 1 mL media containing: no drug treatment; 1 µM MSC2302711A (olaparib); 1 µM AZD2281; or 1 µM both compounds. Nucleofected cells were grown for 3 days before harvest.

[0219] Genomic DNA was harvested from transfected cells using Accutase (MilliporeSigma) and the harvested cells were resuspended in 30 µL QuickExtract reagent (Lucigen, Teddington, Middlesex, UK). Suspensions were incubated at 60 ℃ for 15 m and 95 ℃ for 15 m. Genomic regions targeted by the CBE were amplified by PCR using JumpStart Taq ReadyMix (MilliporeSigma) and the following cycling conditions: 94 ℃/2m; 25 cycles of 94 ℃/30s, 62 ℃/30s, 72 ℃/45s; 72 ℃/5m. Primers are listed in Table 1. PCR products underwent a second round of amplification using Illumina index primers and JumpStart Taq ReadyMix and the following conditions: 95 °C/3m; 9 cycles of 95 ℃/30s, 55 ℃/30s, 72 ℃/30s; 72 ℃/5m. Indexed PCR products were quantified by PicoGreen (ThermoFisher, Waltham, MA), pooled according to DNA content, and purified by Select-a-Size DNA Clean & Concentrator MagBeads (Zymo, Irvine, CA), using 1.2x beads by volume. Pools were diluted to 4 nM. Sequencing was performed on

an Illumina MiSeq instrument using a 300-cycle kit to obtain single-end reads. FASTQ files for each sample were analyzed using a custom analysis script.

[0220] Results are presented in Fig. 4 and Table 2. Table 2 lists the percent increase in the rate of C-to-T editing at two independent target sites, HBB (hemoglobin subunit β) and EMX1-11 (transcription factor) following treatment with each test compound individually and in combination. The data demonstrates that the effect of inhibition of individual DNA repair pathways varies by target site; however, the combination of both SSBR and NHEJ inhibition consistently results in a greater increase in C-to-T editing.

Table 2. Percent increase in the rate of C-to-T editing compared to untreated cells

	EMX1-11	HBB
SSBR inhibition	4.67%	4.88%
NHEJ inhibition	16.51%	1.81%
SSBR inhibition + NHEJ inhibition	21.81%	11.03%

[0221] In Fig. 4, the percent (%) of reads containing indel are plotted following base editing in cells, with and without DNA repair inhibition, individually and in combination. Values are average ± standard deviation for two replicates. All transfections were performed with the same cells, on the same day. The results show that the effect of inhibition of individual DNA repair pathways varies by target site; however, the combination of both SSBR and NHEJ inhibition consistently results in a substantial decrease in the rate of indel formation.

[0222] **Example 2**

[0223] **Simultaneous inhibition of SSBR and NHEJ increases the rate of C-to-T editing and reduces the rate of indel when**

administered in combination with BER inhibition and/or the RNP transfection enhancer dextran sulfate

[0224] SSBR and NHEJ inhibitors were obtained as in Example 1. Cytosine base editor proteins were constructed and purified as in Example 1. Synthetic single guide RNAs (sgRNAs) targeting the genomic site EMX1-11 were purchased from MilliporeSigma. The spacer sequence is given in Table 1. A recombinant uracil glycosylase inhibitor (UGI) containing a Bacillus phage UGI, a c-MYC nuclear localization signal (NLS), and a SV40 Large T antigen NLS (SEQ ID NO: 2) was purified from *E. coli*. Dextran sulfate sodium salt with average molecular weight greater than 500 kDa (Product number: D8906) was purchased from MilliporeSigma. A dextran sulfate (DS) solution was prepared by dissolving the chemical in water at 50 µg/µL and sterilized by filtration through a 0.22 um filter. The stock solution was diluted with water to prepare working solutions of 1 µg/µL.

[0225] Ribonucleoprotein (RNP) complexes were prepared by incubating for 15 minutes at room temperature: 40 pmol of CBE protein, 128 pmol of sgRNA, and 0 or 15 µg of UGI in buffer (20 mM HEPES, 100 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, pH 7.5) for a final volume of 10 µL. RNPs were kept on ice until transfection. HEK293 cells were cultured as in Example 1. At the time of transfection, cells were trypsinized to obtain a single-cell suspension, washed twice with Hank's Balanced Salt Solution and resuspended in Nucleofector Solution V (Lonza) at $5x10^5$ cells per 100 µL. Where indicated, DS solution was added to the cell suspension to a final concentration of 0.5 µg per 100 µL and mixed well by swirling. Nucleofection was performed by mixing 100 µL of prepared cell suspension with 10 µL of complexed CBE RNP by pipetting up and down six times before transferring to a cuvette for electroporation using program Q-001 on a Nucleofector 2b machine. Nucleofected cells were immediately transferred to 12-well plates at a density of 10^5 cells per well in 1 mL media containing: no drug treatment; 1 µM AZD2281 (olaparib); 1 µM MSC2528072A; or 1 µM both compounds. Each experimental condition was tested in two technical replicates. Nucleofected cells were grown for 3 days before

harvest. Genomic DNA was harvested, genomic target sites were amplified, and libraries were prepared and sequenced as described in Example 1.

[0226] Results are presented in Table 3 and Fig. 5. Table 3 gives the percent increase in the rate of C-to-T editing at the EMX1-11 target site following treatment with the test compound individually and in combination, for each transfection condition (RNP alone, with UGI protein, with DS, or with both UGI protein and DS in combination). The data demonstrates that, in each case, the combination of both SSBR and NHEJ inhibition results in an increase in the percentage of C-to-T editing.

Table 3. Percent increase in the rate of C-to-T editing compared to untreated cells

	RNP only	RNP + UGI	RNP + DS	RNP + UGI + DS
SSBR inhibition	3.33%	9.58%	5.38%	5.78%
NHEJ inhibition	4.17%	7.86%	17.20%	11.91%
SSBR inhibition + NHEJ inhibition	10.42%	16.46%	16.94%	17.16%

[0227] In Fig. 5, the % of reads containing indel is plotted following base editing in cells, with and without DNA repair inhibition, individually and in combination, for each transfection condition. Values are average ± standard deviation for two replicates. All transfections were performed with the same cells, on the same day. The results show that, regardless of whether UGI protein or DS were included in the transfection, the combination of SSBR and NHEJ inhibition reliably resulted in a substantial decrease in the rate of indel formation.

[0228] **Example 3**

[0229] **Inhibition of NHEJ Increases the Rate of HDR and Decreases the Rate of Indel Formation**

[0230] The DNA-PK Inhibitors MSC2528072A and MSC2528070A were obtained from Merck Healthcare (Darmstadt, DE) as 10mM solutions in DMSO. See, Fig. 16.

[0231] PURedit Cas9 protein, synthetic single guide RNAs, and single stranded DNA donor templates (single-stranded oligodeoxynucleotides; ssODNs) all were purchased from MilliporeSigma (Burlington, MA).

[0232] Ribonucleoprotein (RNP) complexes were prepared by incubating for 15 minutes at room temperature: 30 pmol of PURedit Cas9 protein and 90 pmol of sgRNA in buffer (20 mM HEPES, 100 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, pH 7.5). 300 pmol of the ssODN was then added for a total volume of 10 µL and then stored on ice until (a prototransfection. K562 cells were obtained from ATCC (Bethesda, MD) and grown at 37 OC and 5% CO2 in Iscove's Modified Dulbecco's Medium supplemented with 10% FBS and GlutaMAX (ThermoFisher). Cells were seeded at 1.00 x 105 cells/mL two days before transfection. At the time of transfection, cells were washed twice with Hank's Balanced Salt Solution and resuspended in Nucleofector Solution SF (Lonza, Walkersville, MD) at 106 cells per 100 µL. Nucleofection was performed by mixing 100 µL of prepared cell suspension with 10 µL of complexed RNP by pipetting up and down six times before transferring to a cuvette for electroporation using program FF-120 on a Nucleofector 4d machine. Nucleofected cells were immediately transferred to cellular media in a conical tube at 4.00 x105 cells/mL. 1.00 x 105 cells were then distributed to wells in a 48-well plate containing media supplemented with DNA-PK inhibitors at various concentrations to obtain wells containing 500 µL of media with 0.25 µM, 0.5 µM, 1.0 µM, 2.0 µM, and 4.0 µM of DNA-PK inhibitors. Some wells had no compounds for controls. Nucleofected cells were grown for 5 days before analysis and harvest. Guide Sequence, Donor Sequence, and NGS Primers found in Fig. 7.

[0233] Cell viability was tested with the CellTiter-Glo® 2.0 Assay (Promega, Madison, WI). An opaque-walled 96-well plate was prepared with 50 µL of each well from the 48-plate after thoroughly mixing and left at room temperature for 30 minutes. 50 µL of room temperature CellTiter-Glo® 2.0 Reagent was then added to each well and mixed on an orbital shaker for 10 minutes. Luminescence was then read on a SpectraMax iD3 (Danaher, Washington, DC).

[0234] Results are presented in Figs. 8 – 11. The results show that the SMI (small molecule inhibitors) compounds did not have an effect on toxicity based on the GLO number given, with no significant difference between transfected cells without a compound added and all 5 doses of both SMI compounds for Cas9 RNP nucleofections in K562 cells targeting BCL4 (a transcriptional co-activator and proto-oncogene candidate) and with no significant difference between all 5 doses of both SMI compounds for Cas9 RNP nucleofections in K562 cells targeting AAVS1 (adeno-associated virus integration site 1) except for at the highest dosage.

[0235] Genomic DNA was harvested from the remaining cells using a MagMAX DNA Multi-Sample Ultra 2.0 Kit (ThermoFisher, Waltham, MA) following manufacturer protocol on a KingFisher Flex (96 deep-well format; ThermoFisher). Genomic regions targeted by the RNP were amplified by PCR using NEBNext® Ultra™ II Q5® (NEB, Ipswich, MA) and the following cycling conditions: 98 OC/30 s; 30 cycles of 98 OC/10 s, 62 OC/30 s, 65 OC/45 s; 65 OC/5 m. Primers are listed in Table 1. PCR products underwent a second round of amplification using Illumina index primers (Illumina, San Diego, CA) and NEBNext® Ultra™ II Q5® and the following conditions: 95 OC/3m; 9 cycles of 95 OC /30 s, 55 OC /30 s, 72 OC C/30 s; 72 OC /5 m. Indexed PCR products were purified by Select-a-Size DNA Clean & Concentrator MagBeads (Zymo Research, Irvine, CA), using 1.2x beads by volume, quantified by PicoGreen (ThermoFisher), and pooled according to DNA content. Pools were diluted to 4 nM. Sequencing was performed on an Illumina MiSeq instrument (Illumina) using a 300-cycle kit to obtain single-end reads. FASTQ files for each sample were analyzed using

Geneious software (GraphPad Software, LLC, Boston, MA), each file had at least 10,000 reads to be included in the data.

[0236] Results are presented in Figs. 12 – 15. The results show the percentage of the three main editing consequences, WT (no changes), HDR (insertion), and NHEJ (random indels) for each condition. For each SMI compound, adding it at the lowest dosage reduces NHEJ and increases HDR, and the trend increases as you increase the dosage until it plateaus at around 60% HDR and 40% NHEJ for the AAVS1 target in K562 cells, and at around 40% HDR and 25% NHEJ for the BCL4 target in K562 cells.

Claims

We Claim:

1. A method for improving the rate of nucleobase conversion and/or reducing the rate of indel formation during CRISPR base editing, said method comprising:
   providing: i) a CRISPR base editor complex comprising: a) a catalytic effector domain, b) a catalytically modified Cas protein, and c) a guide RNA; to form the base editor complex; ii) target cells; and iii) at least one inhibitor selected from one or more of single-strand break repair (SSBR) inhibitors, and nonhomologous end joining (NHEJ) inhibitors;
   introducing said base editor complex into target cells to form transfected cells, and;
   contacting said transfected cells with said one or more inhibitors of SSBR and/or NHEJ to form transfected and treated cells;
   wherein, said transfected and treated cells exhibit one or more of an increase in the rate of nucleobase conversion or decrease in the rate of indel formation when compared to transfected but untreated cells.

2. The method of Claim 1, wherein said inhibitor is an inhibitor of SSBR.

3. The method of Claim 1, wherein said inhibitor is an inhibitor of NHEJ.

4. The method of Claim 1, wherein said transfected cells are treated with both an inhibitor of SSBR and an inhibitor of NHEJ.

5. The method of any of Claims 1 - 4, wherein said transfected and treated cells exhibit an increase in nucleobase conversion.

6. The method of any of Claims 1 - 4, wherein said transfected and treated cells exhibit a decrease in indel formation.

7. The method of Claim 5, wherein said increase in nucleobase conversion is from 2% to 100% when compared to untreated cells.

8. The method of Claim 5, wherein said increase in nucleobase conversion is from 4% to 50% when compared to untreated cells.

9. The method of Claim 5, wherein the nucleobase conversion is from cytosine (C) to thymine (T).

10. The method of Claim 5, wherein the nucleobase conversion is from adenine (A) to guanine (G).

11. The method of Claim 5, wherein the nucleobase conversion is from guanine (G) to cytosine (C) or thymine (T).

12. The method of Claim 6, wherein said decrease in indel formation is from 2% to 90% when compared to untreated cells.

13. The method of Claim 6, wherein said decrease in indel formation is from 5% to 50% when compared to untreated cells.

14. The method of any of Claims 1 - 2 and 4 - 13, wherein said inhibitor of SSBR is an inhibitor of PARP and is selected from a group consisting of Olaparib (AZD2281), Veliparib (ABT-888) and A-966492.

15. The method of any of Claims 1 and 3 - 13, wherein said inhibitor of NHEJ is an inhibitor of DNA-PK and is selected from a group consisting of NU7441, AZD7648, CC-115 and M3814.

16. The method of any of Claims 1 - 15, further comprising a uracil glycosylase inhibitor (UGI) or a nucleotide sequence encoding UGI in the transfection mixture.

17. The method of any of Claims 1 - 16, further comprising dextran sulfate in the transfection mixture.

18. The method of any of Claims 1 - 17, wherein said CRISPR base editing complex is encoded in one or more nucleic acid sequences.

19. The method of Claim 14, wherein said SSBR inhibitor is the PARP1/2 inhibitor MSC2302711A (Olaparib).

20. The method of Claim 19, wherein said inhibitor is at a concentration of 0.1 μM to 10 μM.

21. The method of Claim 15, wherein said NHEJ inhibitor is the DNA-dependent protein kinase inhibitor MSC2528072A and/or MSC2528070A.

22. The method of Claim 21, wherein said inhibitor is at a concentration of 0.1 μM to 10 μM.

23. The method of any of Claims 1 - 22, wherein said comparison of transfected and treated cells to transfected and untreated cells is based on historic data of analogous transfected and untreated cells.

24. The method of any of Claims 1 - 23, wherein said base editing is performed in a eukaryotic cell.

25. The method of Claim 1 - 24, wherein said base editing method is performed *in vivo*.

26. The method of Claim 1 - 24, wherein said base editing method is performed *in vitro*.

27. The method of any of Claims 1 - 26, wherein said method is performed in one or more somatic cells.

28. The method of any of Claims 1 - 26, wherein said method is performed in non-human embryos.

29. The method of any of Claims 1 - 28, wherein the Cas protein is catalytically modified.

30. The method of Claim 29, wherein said catalytically modified Cas protein is a nickase that generates a single strand break on the targeted strand.

31. The method of Claim 30, wherein said catalytically modified Cas protein is catalytically inactive.

32. The method of any of Claims 1 – 31, wherein said base editor complex is introduced into target cells to form transfected cells by a method selected from the group consisting of: electroporation, microinjection, lipofection, calcium phosphate-mediated transfection, nucleofection, cationic polymer transfection, viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection and magnetofection.

33. A method for reducing the rate of indel formation and/or increasing gene editing during gene editing by homology-directed repair (HDR), said method comprising:

providing: i) a site-specific or RNA-guided endonuclease; ii) a donor nucleotide sequence, iii) target cells; and iv) at least one nonhomologous end joining (NHEJ) inhibitor;

introducing said endonuclease and donor nucleotide sequence into target cells to form transfected cells, and;

contacting said transfected cells with said one or more inhibitors of NHEJ to form transfected and treated cells;

wherein, said transfected and treated cells exhibit a decrease in indel formation and/or an increase in HDR when compared to transfected but untreated cells.

34. The method of Claim 33, in which the endonuclease is selected from a zinc-finger nuclease, a TALEN, a meganuclease, a Fanzor nuclease, or a CRISPR complex comprising a Cas protein and a guide RNA.

35. The method of Claim 33 – 34, wherein said transfected and treated cells exhibit an increase in HDR when compared to transfected but untreated cells.

36. The method of Claim 33 – 34, wherein said transfected and treated cells exhibit a decrease in indel formation when compared to transfected but untreated cells.

37. The method of Claims 33 – 35, wherein said increase in HDR is from 2% to 100% when compared to transfected but untreated cells.

38. The method of Claims 33 – 35, wherein said increase in HDR is from 4% to 50% when compared to transfected but untreated cells.

39. The method of Claims 33 – 35, wherein said increase in HDR is over 2% or from 2% to 1000% when compared to transfected but untreated cells.

40. The method of Claims 33 – 34 and 36, wherein said transfected and treated cells exhibit a decrease in indel formation of at least 2% when compared to transfected but untreated cells.

41. The method of Claims 33 – 34 and 36, wherein said decrease in indel formation is from 5% to 50% when compared to transfected but untreated cells.

42. The method of Claims 33 – 34 and 36, wherein said decrease in indel formation is from 2% to 90% or 2% to 100% when compared to transfected but untreated cells.

43. The method of any of Claims 33 – 42, wherein said inhibitor of NHEJ is an inhibitor of DNA-PK and is selected from a group consisting of NU7441, AZD7648, CC-115 and M3814.

44. The method of any of Claims 33 – 42, wherein said NHEJ inhibitor is the DNA-dependent protein kinase inhibitor MSC2528072A and/or MSC2528070A.

45. The method of any of Claims 33 – 44, wherein said inhibitor is at a concentration of 0.1 μM to 10 μM.

46. The method of any of Claims 33 – 45, wherein said comparison of transfected and treated cells to transfected but untreated cells is based on historic data of analogous transfected but untreated cells.

47. The method of any of Claims 33 – 46, wherein said gene editing is performed in a eukaryotic cell.

48. The method of Claims 33 – 47, wherein said gene editing method is performed *in vivo*.

49. The method of Claims 35 – 48, wherein said gene editing method is performed *in vitro*.

50. The method of any of Claims 33 – 49, wherein said gene editing method is performed in one or more somatic cells.

51. The method of any of Claims 33 – 49, wherein said gene editing method is performed in non-human embryos.

52. The method of any of Claims 33 – 50, wherein the endonuclease has full endonuclease activity.

53. The method of any of Claims 33 – 51, wherein the endonuclease is catalytically modified.

54. The method of Claim 53, wherein said catalytically modified endonuclease is a Cas protein and said Cas protein is a nickase that generates a single strand break on the targeted strand.

55. The method of any of Claims 33– 54, wherein said base endonuclease is introduced into target cells to form transfected cells by a method selected from the group consisting of: electroporation, microinjection, lipofection, calcium phosphate-mediated transfection, nucleofection, cationic polymer transfection, viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection and magnetofection.

56. The method of any of Claims 33 - 55, wherein said endonuclease is encoded in one or more nucleic acid sequences.