

DIRECT REPLACEMENT GENOME EDITING**CROSS-REFERENCE**

[001] This application claims the benefit of US Provisional Application Serial Number 63/278,886 filed on November 12, 2021, and US Provisional Application Serial Number 63/341,200 filed on May 12, 2022, the entireties of which are hereby incorporated by reference.

SEQUENCE LISTING

[002] The instant application contains a Sequence Listing which has been submitted electronically in XML file format and is hereby incorporated by reference in its entirety. Said XML copy, created on November 9, 2022, is named Replace Therapeutic 62942-701601 (PCT).xml and is 703,891 bytes in size.”

BACKGROUND

[003] Improved gene editing methods are needed for modifying nucleic acids.

SUMMARY

[004] Disclosed herein, in some aspects, are systems or compositions comprising: a DNA-binding protein coupled to a DNA ligase. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the coupling is covalent. Some aspects include a fusion protein comprising the DNA-binding protein (e.g. endonuclease such as an RNA-guided endonuclease) and the DNA ligase. Some aspects include a composition comprising: a cell containing a DNA-binding protein (e.g. endonuclease such as an RNA-guided endonuclease) and a DNA ligase, both of which are heterologous to the cell. In some aspects, the DNA-binding protein is amino (N)-terminal relative to the DNA ligase within the fusion protein. In some aspects, the DNA-binding protein is carboxy (C)-terminal relative to the DNA ligase within the fusion protein. In some aspects, the connection comprises a linker comprising 1-100 amino acids. In some aspects, the coupling is non-covalent. In some aspects, the composition comprises a first polypeptide comprising at least part of the DNA-binding protein, and a second polypeptide comprising at least part of the DNA ligase, wherein the first and second polypeptides are non-covalently coupled. In some aspects, the first polypeptide comprises a first heterodimerization domain that binds a second heterodimerization domain, and wherein the second polypeptide comprises the second heterodimerization domain. In some aspects, the heterodimer domains comprise a leucine zipper, PDZ domain, streptavidin, streptavidin binding protein, foldon domain, hydrophobic moiety, or a functional binding fragment thereof. In some aspects, the first polypeptide comprises a first intein that binds a second intein, and wherein the second polypeptide comprises the second intein. In some aspects, the ligase comprises a hairpin binding motif, and wherein the DNA-binding protein and the DNA ligase are coupled with a nucleic acid

comprising a scaffold that binds to the DNA-binding protein and a hairpin that binds to the hairpin binding motif. In some aspects, the hairpin binding motif comprises an MS2 coat protein (MCP) peptide, and wherein the hairpin comprises an MS2 hairpin. In some aspects, the DNA-binding protein and the DNA ligase are coupled with a heterobifunctional molecule comprising an endonuclease binding domain and a DNA ligase binding domain. In some aspects, the heterobifunctional molecule comprises a small molecule. In some aspects, the DNA-binding protein comprises a class II CRISPR/Cas endonuclease. In some aspects, the DNA-binding protein comprises a Cas9 endonuclease. In some aspects, the DNA-binding protein comprises a nickase. In some aspects, the DNA-binding protein comprises an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 1-13, or a functional fragment thereof. In some aspects, the DNA ligase ligates DNA strands base paired to a DNA splint. In some aspects, the DNA ligase ligates DNA strands base paired to an RNA splint. In some aspects, the DNA ligase comprises an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 55-96, or a functional fragment thereof. In some aspects, the DNA-binding protein or the DNA ligase comprises a nuclear localization signal, chromatin modifying domain, cell penetrating peptide, or tag polypeptide. Some aspects include a guide RNA and an integrating nucleic acid. Some aspects include one or more nucleic acids encoding the composition. Some aspects include a cell comprising the composition, or comprising the one or more nucleic acids.

[005] Disclosed herein, in some aspects, are editing methods, comprising: contacting a target nucleic acid in a cell with an endonuclease at a predetermined locus of the target nucleic acid, thereby introducing a nick at the predetermined locus of the target nucleic acid; introducing a pre-synthesized integrating nucleic acid to the cell; and ligating a 5' end of the pre-synthesized integrating nucleic acid to a 3' end of the nick at the predetermined locus of the target nucleic acid. In some aspects, the endonuclease comprises a class II CRISPR/Cas endonuclease. In some aspects, the endonuclease comprises Cas9 nickase. Some aspects include contacting the endonuclease and the predetermined locus of the target nucleic acid with a guide nucleic acid. In some aspects, said ligating is performed by a ligase coupled to the endonuclease. In some aspects, the pre-synthesized integrating nucleic acid comprises a mutation in relation to the target nucleic acid. In some aspects, the nick comprises a single phosphodiester strand break in the otherwise double stranded target nucleic acid. In some aspects, the nick comprises a non-sticky, non-blunt end of a strand of the target nucleic acid. In some aspects, the target nucleic acid comprises a chromosome of the cell. In some aspects, the cell is eukaryotic.

[006] Disclosed herein, in some aspects, are editing systems, comprising: a ligase; an endonuclease that introduces a nick at a predetermined locus of a target nucleic acid; and a pre-

synthesized integrating nucleic acid comprising a 5' end that is ligated by the ligase to a 3' end of the nick at the predetermined locus of the target nucleic acid. In some aspects, the endonuclease comprises a class II CRISPR/Cas endonuclease. In some aspects, the endonuclease comprises Cas9 nickase. Some aspects include a guide nucleic acid that brings the endonuclease into proximity with the predetermined locus of the target nucleic acid. In some aspects, the ligase is coupled to the endonuclease. In some aspects, the pre-synthesized integrating nucleic acid comprises a mutation in relation to the target nucleic acid. In some aspects, the nick comprises a single phosphodiester strand break in the otherwise double stranded target nucleic acid. In some aspects, the nick comprises a non-sticky, non-blunt end of a strand of the target nucleic acid. In some aspects, the target nucleic acid comprises a chromosome of a cell. In some aspects, the cell is eukaryotic.

[007] Disclosed herein, in some aspects, are systems of nucleic acids comprising: a guide nucleic acid comprising: (a) a spacer complementary to a region of a genomic locus of a genomic strand, (b) a scaffold for complexing with a DNA-binding protein, (c) an optional donor binding site that is at least partially complementary to an integrating nucleic acid, and (d) a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus; and an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by a DNA-binding protein. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. Disclosed herein, in some aspects, are systems of nucleic acids comprising: a guide nucleic acid comprising: (a) a spacer complementary to a region of a genomic locus of a genomic strand, (b) a scaffold for complexing with a DNA-binding protein, and (c) an optional donor binding site that is at least partially complementary to a splinting nucleic acid; an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by a DNA-binding protein; and a splinting nucleic acid comprising a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus, and comprising an optional guide binding site that is at least partially complementary to a guide nucleic acid. In some aspects, the genomic strand is in a cell. In some aspects, the splinting nucleic acid further comprises a donor binding site that is at least partially identical or complementary to a portion of the integrating nucleic acid. In some aspects, the guide nucleic acid comprises a sequence of linking nucleic acids between the scaffold and the donor binding site. In some aspects, the guide nucleic acid or the integrating nucleic acid comprises a modified internucleoside linkage. In some aspects, the modified internucleoside linkage comprises a phosphorothioate linkage. In some aspects, the modified internucleoside linkage is between any of the 4 terminal nucleosides at a 5' end or at a 3' end of the guide

nucleic acid or the integrating nucleic acid. In some aspects, the guide nucleic acid or the integrating nucleic acid comprises a modified nucleoside. In some aspects, the modified nucleoside comprises a locked nucleic acid (LNA), a 2' fluoro, a 2' O-alkyl, or a combination thereof. In some aspects, the modified nucleoside is any of the 3 terminal nucleosides at a 5' end or at a 3' end of the guide nucleic acid or the integrating nucleic acid. The modified nucleoside may include an LNA, a 2' fluoro, a 2' O-alkyl, a methylated cytosine, an inverted thymidine, or a combination thereof.

[008] Disclosed herein, in some aspects, are compositions, comprising: a DNA-binding protein connected to a DNA ligase. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the connection between the DNA-binding protein and the DNA ligase is covalent. Some aspects include a fusion protein comprising the DNA-binding protein upstream of the DNA ligase. Some aspects include a fusion protein comprising the DNA-binding protein downstream of the DNA ligase. In some aspects, the connection comprises a linker comprising 1-100 amino acids. In some aspects, the composition comprises a first polypeptide comprising at least part of the DNA-binding protein, and a second polypeptide comprising at least part of the DNA ligase, wherein the first and second polypeptides are bound together covalently or non-covalently. In some aspects, the first polypeptide comprises a first heterodimerization domain that binds a second heterodimerization domain, and wherein the second polypeptide comprises the second heterodimerization domain. In some aspects, the heterodimer domains comprise a leucine zipper, PDZ domain, streptavidin, streptavidin binding protein, foldon domain, hydrophobic moiety, or a functional binding fragment thereof. In some aspects, the first polypeptide comprises a first intein that binds a second intein, and wherein the second polypeptide comprises the second intein. In some aspects, the DNA-binding protein and the DNA ligase are bound together by a small molecule. In some aspects, the DNA-binding protein comprises a class II CRISPR/Cas endonuclease. In some aspects, the DNA-binding protein comprises a Cas9 endonuclease. In some aspects, the DNA-binding protein comprises a nickase. In some aspects, the DNA-binding protein comprises an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 1-13, or a functional fragment thereof. In some aspects, the DNA ligase ligates DNA strands base paired to a DNA splint. In some aspects, the DNA ligase ligates DNA strands base paired to an RNA splint. In some aspects, the DNA ligase comprises an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 55-96, or a functional fragment thereof. In some aspects, the DNA-binding protein or the DNA ligase comprises a nuclear localization signal, chromatin modifying domain, cell penetrating peptide, or tag polypeptide. Some aspects include a guide

RNA and an integrating nucleic acid. Some aspects relate to a cell comprising the composition. Some aspects include a nucleic acid encoding the composition. Some aspects include one or more nucleic acids encoding the first or second polypeptides. Some aspects include an editing method (e.g. nucleic acid) which uses the composition. Some aspects include a method of treatment using the composition. Some aspects include administering the composition to a subject.

[009] Disclosed herein, in some aspects, are fusion proteins, comprising: a DNA-binding protein fused to a DNA ligase. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. Disclosed herein, in some aspects, are protein complexes, comprising: a DNA-binding protein bound to a DNA ligase. In some aspects, the endonuclease and the DNA ligase are bound together through heterodimerization domains. In some aspects, the heterodimerization domains comprise leucine zippers, PDZ domains, streptavidin and streptavidin binding protein, foldon domains, hydrophobic polypeptides, an antibody that binds the Cas nickase, or an antibody that binds the DNA ligase, or one or more binding fragments thereof. Disclosed herein, in some aspects, are cells comprising the fusion protein or the protein complex. Disclosed herein, in some aspects, are cells comprising a heterologous DNA-binding protein and a DNA ligase that was introduced into the cell. Some aspects include a nuclease that is different from the DNA-binding protein. Disclosed herein, in some aspects, are guide nucleic acids, comprising: a spacer at least partially reverse complementary to a first region of a target nucleic acid; a scaffold configured to bind to an endonuclease; and a flap binding site at least partially reverse complementary to a nucleic acid flap, and an integrating nucleic acid binding site. Disclosed herein, in some aspects, are integrating nucleic acids, comprising: a single or double-stranded DNA region to be inserted into a target nucleic acid, wherein the single or double-stranded DNA region is flanked by at least one additional single-stranded region comprising a guide binding site. Disclosed herein, in some aspects, are editing systems, comprising a DNA-binding protein, the guide nucleic acid, and the integrating nucleic acid. Disclosed herein, in some aspects, are editing methods, comprising: contacting a target nucleic acid with the editing system and a DNA ligase.

[0010] Disclosed herein, in some aspects, are systems comprising: at least one DNA-binding protein; at least one guide nucleic acid comprising: a spacer at least partially complementary to a genomic locus in a cell; a scaffold for complexing with the at least one DNA-binding protein; and an optional donor binding site that is at least partially complementary to an integrating nucleic acid; and at least one DNA ligase; and the integrating nucleic acid, comprising a flap binding site at least partially reverse complementary to a nucleic acid flap and optionally comprising a guide binding site that is at least partially complementary to the at least one guide

nucleic acid, wherein the at least one DNA-binding protein cleaves or nicks at least one strand of the genomic locus, and wherein the at least one DNA ligase ligates an end of the integrating nucleic acid to the genomic flap site, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the integrating nucleic acid comprises a single-stranded DNA. In some aspects, the integrating nucleic acid comprises a double-stranded DNA.

[0011] Disclosed herein, in some aspects, are systems comprising: at least one DNA-binding protein comprising a first DNA-binding protein and an optional second DNA-binding protein; at least one guide nucleic acid comprising a first guide nucleic acid and a second guide nucleic acid, the first guide nucleic acid comprising: a first spacer complementary to a first region of a genomic locus in a cell; a first scaffold for complexing with the first DNA-binding protein; and an optional first donor binding site that at least partially complementary to an integrating nucleic acid; and a first flap binding site that is at least partially identical or complementary to a first genomic flap at or adjacent to the genomic locus; and the second guide nucleic acid comprising: a second spacer complementary to a second region of the genomic locus in the cell; a second scaffold for complexing with the first or second DNA-binding protein; an optional second donor binding site that at least partially complementary to the integrating nucleic acid; and a second flap binding site that is at least partially identical or complementary to a second genomic flap at or adjacent to the genomic locus; at least one DNA ligase comprising a first DNA ligase and an optional second DNA ligase; and at least one integrating nucleic acid comprising a first strand and a second strand: wherein the first strand comprises an optional first guide binding site that is at least partially complementary to the first guide nucleic acid; and wherein the second strand comprises an optional second guide binding site that is at least partially complementary to the second guide nucleic acid, wherein the first DNA-binding protein and/or the second DNA-binding protein each cleaves or nicks at least one strand of the genomic locus in the cell; and wherein the first DNA ligase ligates an end of the first strand of the integrating nucleic acid to the first genomic flap; and the first or second DNA ligase ligates an end of the second strand of the integrating nucleic acid to the second genomic flap, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. In some aspects, the integrating nucleic acid comprises a double-stranded DNA duplex region. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the integrating nucleic acid comprises a 5' overhang optionally comprising the first guide binding site. In some aspects, the integrating nucleic acid comprises a 5' overhang optionally comprising the second guide binding site.

[0012] Disclosed herein, in some aspects, are systems comprising: at least one DNA-binding protein; at least one guide nucleic acid comprising: a spacer complementary to a genomic locus in a cell; a scaffold for complexing with the at least one DNA-binding protein; and an optional donor binding site that is at least partially complementary to an integrating nucleic acid; at least one DNA ligase; and the integrating nucleic acid that: comprises an optional guide binding site that is at least partially complementary to the at least one guide nucleic acid; and comprises a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus, wherein the at least one DNA-binding protein cleaves or nicks at least one strand of the genomic locus; and wherein the at least one DNA ligase ligates an end of the integrating nucleic acid to the genomic flap, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the integrating nucleic acid comprises a DNA comprising a 3' overhang. In some aspects, the 3' overhang comprises the guide binding site. In some aspects, the 3' overhang comprises the flap binding site. In some aspects, the at least one DNA ligase ligates a strand of the integrating nucleic acid to the genomic nucleic acid sequence.

[0013] Disclosed herein, in some aspects, are systems comprising: at least one DNA-binding protein comprising a first DNA-binding protein and an optional second DNA-binding protein; at least one guide nucleic acid comprising a first guide nucleic acid and a second guide nucleic acid, the first guide nucleic acid comprising: a first spacer complementary to a first region of a genomic locus in a cell; a first scaffold for complexing with the first DNA-binding protein; and an optional first donor binding site that at least partially complementary to an integrating nucleic acid; and the second guide nucleic acid comprising: a second spacer complementary to a second region of the genomic locus in the cell; a second scaffold for complexing with the first or second DNA-binding protein; and an optional second donor binding site that at least partially complementary to the integrating nucleic acid; and at least one DNA ligase comprising a first DNA ligase and an optional second DNA ligase; and the integrating nucleic acid comprising a first strand and a second strand: wherein the first strand comprises an optional first guide binding site that is at least partially complementary to the first guide nucleic acid; wherein the second strand comprises an optional second guide binding site that is at least partially complementary to the second guide nucleic acid; wherein the first strand comprises a first flap binding site that is at least partially identical or complementary to a first genomic flap at or adjacent to the genomic locus; and wherein the second strand comprises a second flap binding site that is at least partially identical or complementary to a second genomic flap at or adjacent to the genomic locus; wherein the first DNA-binding protein and/or the second DNA-binding

protein each cleaves or nicks at least one strand of the genomic locus in the cell; and wherein the first DNA ligase ligates an end of the first strand of the integrating nucleic acid to the first genomic flap; and the first or second DNA ligase ligates an end of the second strand of the integrating nucleic acid to the second genomic flap, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the integrating nucleic acid comprises a double-stranded DNA duplex region. In some aspects, the double-stranded DNA comprises a 3' overhang optionally comprising the first guide binding site, and comprising the first flap binding site. In some aspects, the double stranded DNA comprises a 3' overhang optionally comprising the second guide binding site, and comprising the second flap binding site.

[0014] The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the at least one DNA-binding protein comprises a Cas protein or a functional fragment thereof. In some aspects, the Cas protein or the functional fragment thereof comprises nickase activity. In some aspects, the at least one DNA-binding protein comprises a Cas9 nickase or a functional fragment thereof. In some aspects, the at least one DNA ligase ligates nucleic acids bound to DNA. In some aspects, the at least one DNA ligase ligates nucleic acids bound to RNA. In some aspects, the at least one DNA ligase comprises a PBCV-1 DNA ligase. In some aspects, the at least one DNA ligase is operatively coupled to the at least one DNA-binding protein. In some aspects, the at least one DNA ligase is fused to the at least one DNA-binding protein as a fusion polypeptide. In some aspects, the at least one DNA-binding protein and the at least one DNA ligase each comprises a heterodimer domain. In some aspects, the at least one DNA-binding protein and the at least one DNA ligase forms a heterodimer via the heterodimer domain. In some aspects, the at least one DNA-binding protein comprises a linker. In some aspects, the linker connects the Cas protein or a functional fragment thereof to the heterodimer domain. In some aspects, the at least one DNA-binding protein comprises a localization signal sequence. In some aspects, the at least one DNA ligase comprises a localization signal sequence. In some aspects, the localization signal sequence comprises a nuclear localization sequence (NLS). In some aspects, the at least one DNA-binding protein or the at least one DNA ligase are directed to nucleus of the cell by the NLS. In some aspects, the at least one integrating nucleic acid corrects at least one genetic mutation in the at least one genomic locus. In some aspects, the at least one integrating nucleic acid inserts a coding sequence. In some aspects, the coding sequence encodes a full length protein. In some aspects, the at least one integrating nucleic acid inserts a non-coding sequence. In some aspects, the non-coding sequence knocks out an endogenous gene. In some aspects, the non-coding

sequence comprises a regulatory element. Some aspects further include a nuclease. In some aspects, the nuclease comprises an exonuclease for digesting the genomic flap. In some aspects, the nuclease comprises a human flap endonuclease 1 (hFEN1), a human exonuclease 5 (hEXO5), a T5 exonuclease, a T7 exonuclease, an exonuclease VIII, a flap endonuclease domain of *E. coli* PolI, a RecJF, a Lambda exonuclease, a Xni (ExoIXI), a SaFEN (*Staphylococcus aureus* FEN), a nuclease BAL-31, or a fragment thereof. In some aspects, the heterologous nuclease comprises an endonuclease for digesting the genomic flap, and the endonuclease is different from the at least one DNA-binding protein. In some aspects, the at least one DNA-binding protein comprises at least one additional functional domain. In some aspects, the at least one additional functional domain comprises a chromatin modifying domain. In some aspects, the at least one additional functional domain comprises a cell penetrating peptide. In some aspects, the at least one guide nucleic acid comprises at least one nucleic acid modification. In some aspects, the at least one nucleic acid modification comprises a modification to a backbone, a sugar, a base, or a combination thereof. In some aspects, the at least one DNA-binding protein is complexed with the at least one guide nucleic acid. In some aspects, the at least one guide nucleic acid is complexed with the integrating nucleic acid. In some aspects, the at least one DNA-binding protein, the at least one guide nucleic acid, the at least one at least one DNA ligase, the integrating nucleic acid, or a combination thereof is encoded by a polynucleotide. In some aspects, the polynucleotide comprises mRNA. In some aspects, the polynucleotide comprises a vector. In some aspects, the vector comprises a viral vector. In some aspects, the at least one DNA-binding protein, the at least one guide nucleic acid, the at least one at least one DNA ligase, the integrating nucleic acid, or a combination thereof is encapsulated by at least one lipid nanoparticle. In some aspects, the cell comprises a bacterial cell, an eukaryotic cell, or a plant cell. In some aspects, the eukaryotic cell comprises a mammalian cell. Some aspects include a composition comprising the system. Some aspects include a cell comprising the system. Some aspects include a cell line comprising the cell. Some aspects include a pharmaceutical composition comprising the system. Some aspects include a pharmaceutical composition comprising the composition. Some aspects include a pharmaceutical composition comprising the cell. Some aspects include a pharmaceutically acceptable: excipient, carrier, or diluent. In some aspects, the pharmaceutical composition is formulated for administering intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally, intracerebellarly, intracerebroventricularly, intraperenchymally, subcutaneously, intratumorally, pulmonarily, endotracheally, intraperitoneally, intravesically, intravaginally, intrarectally, orally, sublingually, transdermally, by inhalation, by inhaled nebulized form, by intraluminal-GI route, or a combination thereof to

a subject in need thereof. Some aspects include a kit comprising: the system, the composition, or the pharmaceutical composition and a container. In some aspects, include method for modifying a cell comprising contacting a cell with the system. In some aspects, include method for modifying a cell comprising contacting a cell with the composition. In some aspects, include method for modifying a cell comprising contacting a cell with the pharmaceutical composition. In some aspects, the cell is not a dividing cell. In some aspects, the integrating nucleic acid is inserted into the genomic locus of the cell independent of endogenous non-homologous end joining (NHEJ) and independent of endogenous homology-directed repair (HDR). Some aspects include a method for treating a disease or condition in subject in need thereof comprising: contacting the cell or the subject with the system, the composition, or the pharmaceutical composition; replacing a genomic locus in a cell with an integrating nucleic acid, thereby treating the disease or condition in the subject. In some aspects, the cell is not a dividing cell. In some aspects, the integrating nucleic acid is inserted into the genomic locus of the cell independent of endogenous non-homologous end joining (NHEJ) and independent of endogenous homology-directed repair (HDR).

[0015] Disclosed herein, in some aspects, are guide nucleic acids comprising: a spacer that is at least partially complementary to a genomic locus in a cell; a scaffold for complexing with a DNA-binding protein; and a donor binding site that is at least partially complementary to an integrating nucleic acid. The DNA-binding protein may include an endonuclease. The endonuclease may include an RNA-guided endonuclease. In some aspects, the guide nucleic acid comprises a flap binding site that is at least partially complementary to a genomic sequence of the genomic locus. In some aspects, the guide nucleic acid comprises at least one nucleic acid modification. In some aspects, the at least one nucleic acid modification comprises a modification to a backbone, a sugar, a base, or a combination thereof. In some aspects, the guide nucleic acid comprises RNA sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] Fig. 1A illustrates a guide nucleic acid, an endonuclease, a ligase, and a donor strand at a genomic locus.

[0017] Fig. 1B follows sequentially from **Fig. 1A**, and illustrates a donor strand incorporated into one side of a genomic locus, the donor strand having displaced a genomic flap.

[0018] Fig. 1C follows sequentially from **Fig. 1B**, and illustrates a donor strand incorporated into one side of a genomic locus, and a nick appearing where a genomic flap has been removed.

[0019] Fig. 2A illustrates 2 guide nucleic acids, 2 endonucleases, 2 ligases, and a donor strand at a genomic locus.

[0020] **Fig. 2B** follows sequentially from **Fig. 2A**, and illustrates a donor strand incorporated into a genomic locus, the donor strand having displaced 2 genomic flaps.

[0021] **Fig. 2C** follows sequentially from **Fig. 2B**, and illustrates a donor strand incorporated into a genomic locus, and 2 nicks appearing where genomic flaps have been removed.

[0022] **Fig. 3A** illustrates a guide nucleic acid, an endonuclease, a ligase, and a donor strand at a genomic locus.

[0023] **Fig. 3B** follows sequentially from **Fig. 3A**, and illustrates a donor strand incorporated into one side of a genomic locus, the donor strand having displaced a genomic flap.

[0024] **Fig. 3C** follows sequentially from **Fig. 3B**, and illustrates a donor strand incorporated into one side of a genomic locus, and a nick appearing where a genomic flap has been removed.

[0025] **Fig. 4A** illustrates 2 guide nucleic acids, 2 endonucleases, 2 ligases, and a donor strand at a genomic locus.

[0026] **Fig. 4B** follows sequentially from **Fig. 4A**, and illustrates a donor strand incorporated into a genomic locus, the donor strand having displaced 2 genomic flaps.

[0027] **Fig. 4C** follows sequentially from **Fig. 4B**, and illustrates a donor strand incorporated into a genomic locus, and 2 nicks appearing where genomic flaps have been removed.

[0028] **Fig. 5A** illustrates a guide nucleic acid, an endonuclease, a ligase, and a donor strand at a genomic locus.

[0029] **Fig. 5B** follows sequentially from **Fig. 5A**, and illustrates a donor strand incorporated into a genomic locus, the donor strand having displaced a genomic flap.

[0030] **Fig. 5C** follows sequentially from **Fig. 5B**, and illustrates a donor strand incorporated into one side of a genomic locus, and a nick appearing where a genomic flap has been removed.

[0031] **Fig. 6A** illustrates 2 guide nucleic acids, 2 endonucleases, 2 ligases, and a donor strand at a genomic locus.

[0032] **Fig. 6B** follows sequentially from **Fig. 6A**, and illustrates a donor strand incorporated into a genomic locus, the donor strand having displaced 2 genomic flaps.

[0033] **Fig. 6C** follows sequentially from **Fig. 6B**, and illustrates a donor strand incorporated into a genomic locus, and 2 nicks appearing where genomic flaps have been removed.

[0034] **Fig. 7** illustrates some examples of fusion protein arrangements.

[0035] **Fig. 8A** illustrates an exemplary nicking and ligation pattern of an integrating nucleic acid.

[0036] **Fig. 8B** illustrates a DNA gel showing a pattern associated with 1-Sided Replacer 2 performed *in vitro* using 30nt GBS/DBS and thermostable T4 ligase. Using a 30nt GBS/DBS combination, a donor containing a protospacer adjacent motif (PAM) mutation, and a thermostable T4 ligase (Hi-T4, NEB), we were able to produce a final Replacer product (Lane

3) corresponding to the size of our control product (Lane 1). Replacer products were not detected in the absence of nicking Cas9 (Cas9n) (Lane 2), or in the absence of the bottom donor which serves as the splint (Lanes 4 & 5).

[0037] Fig. 8C illustrates an exemplary nucleic acid gel showing pattern associated with *in vitro* 1-Sided Replacer 2 using variable length GBS/DBS combinations and T4 ligase. Using regular T4 ligase (NEB), we were to produce a final Replacer product corresponding to the size of the control when using multiple GBS/DBS combinations, including no GBS/DBS, 20nt GBS/DBS, and 30nt GBS/DBS. Additionally, in this experiment, recoded dsDNA donors containing PAM mutation were more efficient at producing final Replacer products compared to PAM mutant dsDNA donors that were not recoded.

[0038] Fig. 9 illustrates measurement of a percentage of cells expressing green fluorescent protein (GFP), indicating gene editing from BFP to GFP by a 1-sided Replacer 2 with nicking Cas9 and DNA ligase.

[0039] Fig. 10 illustrates sequencing reads merged and aligned to an amplicon of interest and a percentage of total reads that matched an intended edit via a 1-sided replacer 2 with a nicking Cas9 and a T4 DNA ligase.

[0040] Fig. 11 illustrates sequencing reads merged and aligned to an amplicon of interest and a percentage of total reads that matched an intended edit via a 2-sided replacer 2 with a nicking Cas9 and a T4 DNA ligase.

[0041] Fig. 12 illustrates measurement of a percentage of cells expressing green fluorescent protein (GFP), indicating gene editing from BFP to GFP via a 1-Sided Replacer 2 with a nicking Cas9 and a T4 DNA Ligase.

DETAILED DESCRIPTION

Introduction

[0042] Recent advances in gene editing tools have enabled precision editing of genomes for therapeutic, agricultural, industrial, and research purposes. Some nuclease-based tools such as CRISPR-Cas9 use a guide RNA to target the Cas9 protein to a specific DNA sequence specified by the spacer sequence in the guide RNA. Cas9 nuclease activity then cleaves the DNA resulting in a double-stranded break (DSB). DSBs are typically repaired through endogenous DNA repair mechanisms including non-homologous end joining (NHEJ) or homology-directed repair (HDR). However, NHEJ results in a spectrum of nucleotide insertions and deletions (indels) that hinder its utility for precision editing. HDR efficiency is very low in nondividing cells and may require DNA replication. Even when HDR editing is detectable, DSB-induced indels are often prevalent, meaning that HDR may not be feasible when precision editing is desired.

[0043] Homology-independent targeted insertion (HITI) utilizes NHEJ DNA repair mechanisms active in nondividing cells for CRISPR-guided transgene integration in nondividing cells such as primary neurons, retinal pigment epithelial cells, and HSPCs. However, due to the generation of DSBs from Cas9, HITI generates high frequencies of indels, resulting in unintended mutations in addition to DSB associated toxicity.

[0044] Other methods for gene editing have additional limitations. Tools employing fusions of nicking Cas nucleases with nucleotide deaminases (e.g. base editors) can perform certain nucleotide mutations, e.g. cytosine base editors can convert C to T. While some base editors can perform precision editing at high efficiency, they are inherently limited to specific edits determined by the deaminase variant so they are only applicable to specific substitution mutations and further cannot perform precise insertion or deletion edits. Moreover, base editors are generally limited to a small editing window within a subset of the protospacer region and are therefore significantly limited by protospacer adjacent motif (PAM) availability. Finally, base editors can exhibit bystander mutations within the editing region (e.g. if two C's are present) and have demonstrated DNA and RNA off-target deaminase activity.

[0045] Existing precision editing technologies have limitations that hamper their practical applicability in a variety of ways. In particular, they may rely on endogenous cellular machinery for editing, for example HDR machinery for nuclease-based editing and mismatch repair for base editing. No system has been reported that is independent of all endogenous factors. Reliance on endogenous factors is problematic because different cell types have different activity levels of these endogenous factors, and in many cases the activity is not sufficient to provide useful levels of editing. An example where this reliance is particularly problematic is nondividing cells, which comprise the majority of cells in adults and therefore are not amenable to many existing precision editing tools.

[0046] Accordingly, there remains a need for a system or a method for effective gene editing or for modifying gene expression by gene editing. Particularly, there remains a need for the system or method for gene editing or modifying gene expression, where the system or the method do not rely on the endogenous components or mechanism of a cell. There also remains a need for a system or a method for correcting genetic mutations in a cell. In some cases, the correction of genetic mutation can treat a disease or condition in subject in need thereof. As will be seen below, the systems, methods, and compositions disclosed herein may be useful for addressing these needs or limitations.

Overview

[0047] Described herein are self-contained gene editing systems. In some such self-contained systems, every aspect of gene editing may be controlled. Some such systems do not rely on host

cell machinery to perform an editing function, or to replace or repair any aspect of a target nucleic acid such as a genomic locus. Some such systems are unaffected by a cell's nucleotide triphosphate (dNTP) concentration because the editing may be performed without use of a polymerase. For example, an integrating nucleic acid may be delivered and inserted into a genetic locus without transcribing a template. The editing may exclude a need to rely on a cell repair system such as HDR or NHEJ. The editing may be performed without cell cycling. The gene editing may take place in a cell, or may even be performed *in vitro*. For example, the gene editing may even be performed in a test tube or outside of a cell.

[0048] Described herein are systems and methods for editing DNA with a donor strand without generating a double-stranded break in the genome using CRISPR-guided DNA ligases and guide nucleic acids targeting the genomic region of interest. DNA ligases are enzymes which chemically join two DNA molecules via a phosphodiester bond. DNA ligases may or may not require hybridization of the DNA molecules to a DNA or RNA backbone or “splint” which is reverse complementary to the DNA sequences that are to be ligated. Targeting of ligases to genomic nicks generated by CRISPR nucleases enables precise replacement of genomic DNA with donor strands optionally recruited by guide nucleic acids into targeted loci. The CRISPR-guided DNA ligases can be composed of DNA ligases that are fused, recruited, or unfused to the RNA-guided endonuclease by utilizing peptide linkers, heterodimerization domains, or two separate peptides, respectively.

[0049] Some aspects include a cell containing or comprising an RNA-guided endonuclease and a DNA ligase, both of which are introduced into the cell. The endonuclease or ligase may be heterologous to the cell. The endonuclease and ligase may be heterologous to the cell. The ligase may be endogenous to the cell. In some aspects, a cell comprises an RNA-guided endonuclease and a DNA ligase, both of which are heterologous to the cell. The cell may include a composition or system described herein. The cell may be used or included in a system, composition, or method described herein.

[0050] A system described herein may include a heterologous endonuclease comprising an RNA-guided endonuclease such as nicking Cas9 as well as a heterologous ligase (e.g., a DNA ligase) that can utilize an RNA splint. The guide nucleic acid optionally recruits a donor strand to the site targeted by the endonuclease (e.g., a targeted genomic locus) and also generates a splint across from the donor strand (donor strand) and genomic flap generated by the nicking Cas9, resulting in ligation of the donor strand and the genomic flap by the DNA ligase. In some embodiments, the ligase is or comprises an endogenous ligase. The system can utilize one or more guide nucleic acids that together can comprise the following components, optionally in the following order: 5' spacer - scaffold - donor binding site (optional) - flap binding site 3'.

The donor strand (donor strand) can comprise the following sequence components: 5' guide binding site - donor strand 3'. The guide binding site of the donor strand is at least partially reverse complementary to the donor binding site of the guide nucleic acid such that the donor hybridizes to the guide and is localized to the target site of the RNA guided endonuclease. The 5' end of the donor sequence and the 3' end of the genomic flap generated by nuclease nicking activity are ligated by the DNA ligase, splinted by the donor binding site and a flap binding site of the guide nucleic acid(s).

[0051] Fig. 1A-1C illustrate a non-limiting example of a system (1-sided Replacer 1). The example includes a guide nucleic acid comprising: a spacer for targeting a genomic locus; a scaffold for complexing and recruiting an endonuclease described herein; a donor binding site for complexing with a donor strand; and a flap binding site for complexing with a genomic flap of the genomic locus. The guide nucleic acid is shown complexed with an endonuclease (e.g., a Cas9 nickase, nCas9) operatively coupled to a ligase. The guide nucleic acid may direct the endonuclease to a genomic locus that is bound by the spacer of the guide nucleic acid. The guide nucleic acid is also shown as partially complementary to a donor strand (complexing between the donor binding site of the guide nucleic acid and guide binding site of the donor strand). The endonuclease, when directed by the guide nucleic acid, can cleave or nick at least one strand of the genomic locus, and the ligase can ligate one end of the donor strand with the cleaved or nicked end of the genomic locus, thus incorporating the donor strand into the genomic locus. The incorporation of the donor strand into the genomic locus may generate a genomic flap that can be digested and removed by a nuclease.

[0052] Fig. 2A-2C illustrate a non-limiting example of a system (2-sided Replacer 1). The guide nucleic acid in the example, similar to the guide nucleic acid of **Fig 1A**, comprises: a spacer for targeting a genomic locus; a scaffold for complexing and recruiting an endonuclease described herein; a donor binding site for complexing with a donor strand; and a flap binding site for complexing with a genomic flap of the genomic locus. In **Fig. 2A**, a first guide nucleic acid is shown complexed with a first endonuclease operatively coupled with a first ligase and a second guide nucleic acid is complexed with a second endonuclease operatively coupled with a second ligase. The first endonuclease and the second nuclease may each cleave at least one strand of the genomic locus. The two cleaved ends of the genomic locus can then be ligated to the two ends of the donor strand, thereby incorporating the donor strand into the genomic locus. The insertion of the donor strand at the genomic locus may generate two genomic flaps that can be digested and removed by a nuclease.

[0053] Fig. 3A-3C illustrate a non-limiting example of a system (1-sided Replacer 2). In the example, a guide nucleic acid comprises: a spacer for targeting a genomic locus; a scaffold for

complexing and recruiting an endonuclease described herein; and a donor binding site for complexing with a donor strand. Also shown in **Fig. 3A** is a donor strand comprising at least one overhang, where the overhang comprises: a flap binding site for complexing with a genomic flap of the genomic locus; and a guide binding site for complexing with the guide nucleic acid (via the donor binding site of the guide nucleic acid). The guide nucleic acid can be complexed with an endonuclease (e.g., nCas9) operatively coupled to a ligase. The guide nucleic acid in the example directs the endonuclease and the ligase to a genomic locus that is bound by the spacer of the guide nucleic acid. The guide nucleic acid in the example is also partially complementary to a donor strand (complexing between the donor binding site of the guide nucleic acid and guide binding site of the donor strand). The endonuclease, when directed by the guide nucleic acid, can cleave at least one strand of the genomic locus, and the ligase can ligate one end of the donor strand with the cleaved end of the genomic locus, thus incorporating the donor strand into the genomic locus. The incorporation of the donor strand into the genomic locus may generate a genomic flap that can be digested and removed by a nuclease.

[0054] **Fig. 4A-4C** illustrates a non-limiting example of a system (2-sided Replacer 2). In the example, where the guide nucleic acid, similar to the guide nucleic acid of **Fig 3A**, comprises a spacer for targeting a genomic locus; a scaffold for complexing and recruiting an endonuclease described herein; and a donor binding site for complexing with a donor strand. Also shown in **Fig. 4A** is a donor strand comprising two overhangs, where the overhangs each comprise a flap binding site for complexing with a genomic flap of the genomic locus; and a guide binding site for complexing with a guide nucleic acid (via a donor binding site of the guide nucleic acid). The flap binding site of the donor strand can bring the donor strand in close proximity with the genomic locus after a genomic flap is generated after the endonuclease cleaves at least one strand of the genomic locus. In **Fig. 4A**, a first guide nucleic acid is shown complexed with a first endonuclease operatively coupled with a first ligase and a second guide nucleic acid is complexed with a second endonuclease operatively coupled with a second ligase. In the example, the first endonuclease and the second nuclease each cleave at least one strand of the genomic locus. The two cleaved ends of the genomic locus can then be ligated to the two ends of the donor strand, thereby incorporating the donor strand into the genomic locus. In the example, the insertion of the donor strand at the genomic locus generates two genomic flaps that can be digested and removed by a nuclease.

[0055] A system described herein (Replacer 3) may include a heterologous endonuclease comprising an RNA-guided endonuclease such as nicking Cas9 as well as a ligase (e.g., a DNA ligase) that can utilize a DNA splint. The guide nucleic acid optionally recruits a donor strand to the site targeted by the endonuclease (e.g., a targeted genomic locus) and also generates a splint

across from the donor strand (donor strand) and genomic flap generated by the nicking Cas9, resulting in ligation of the donor strand and the genomic flap by the DNA ligase. At least part of the flap binding site and donor binding site on the guide nucleic acid are DNA such that ligases that utilize DNA splints are able to catalyze the intended reaction. The system can utilize one or more guide nucleic acids that together can comprise the following components, optionally in the following order: 5' spacer - scaffold - donor binding site (optional) - flap binding site 3'. The donor strand (donor strand) can comprise the following sequence components: 5' guide binding site - donor strand 3'. The guide binding site of the donor strand is at least partially reverse complementary to the donor binding site of the guide nucleic acid such that the donor hybridizes to the guide and is localized to the target site of the RNA guided endonuclease. The 5' end of the donor sequence and the 3' end of the genomic flap generated by nuclease nicking activity are ligated by the DNA ligase, splinted by the donor binding site and a flap binding site of the guide nucleic acid(s).

[0056] Fig. 5A-5C illustrate a non-limiting example of a system (1-sided Replacer 3). The example includes a guide nucleic acid comprising: a spacer for targeting a genomic locus; a scaffold for complexing and recruiting an endonuclease described herein; a donor binding site for complexing with a donor strand; and a flap binding site for complexing with a genomic flap of the genomic locus, wherein at least part of the flap binding site and donor binding site are comprised of DNA. The guide nucleic acid is shown complexed with an endonuclease (e.g., a Cas9 nickase, nCas9) operatively coupled to a ligase (e.g., an endogenous ligase or an exogenous ligase). The guide nucleic acid may direct the endonuclease to a genomic locus that is bound by the spacer of the guide nucleic acid. The guide nucleic acid is also shown as partially complementary to a donor strand (complexing between the donor binding site of the guide nucleic acid and guide binding site of the donor strand). The endonuclease, when directed by the guide nucleic acid, can cleave at least one strand of the genomic locus, and the ligase can ligate one end of the donor strand with the cleaved end of the genomic locus, thus incorporating the donor strand into the genomic locus. The incorporation of the donor strand into the genomic locus may generate a genomic flap that can be digested and removed by a nuclease.

[0057] Fig. 6A-6C illustrate a non-limiting example of a system (2-sided Replacer 3). The guide nucleic acid in the example, similar to the guide nucleic acid of **Fig. 5A**, comprises: a spacer for targeting a genomic locus; a scaffold for complexing and recruiting an endonuclease described herein; a donor binding site for complexing with a donor strand; and a flap binding site for complexing with a genomic flap of the genomic locus, wherein at least part of the flap binding site and donor binding site are comprised of DNA. In **Fig. 6A**, a first guide nucleic acid is shown complexed with a first endonuclease operatively coupled with a first ligase and a

second guide nucleic acid is complexed with a second endonuclease operatively coupled with a second ligase. The first endonuclease and the second nuclease may each cleave at least one strand of the genomic locus. The two cleaved ends of the genomic locus can then be ligated to the two ends of the donor strand, thereby incorporating the donor strand into the genomic locus. The insertion of the donor strand at the genomic locus may generate two genomic flaps that can be digested and removed by a nuclease.

[0058] Ligation may be performed using a DNA ligase that can utilize an RNA splint such as SplintR ligase--also known as PBCV-1 DNA Ligase--from *Chlorella* virus. In some aspects, the system utilizes two guide nucleic acids targeting the CRISPR-guided ligase to target sites on opposite strands flanking the genomic region of interest. In some aspects, each guide nucleic acid interacts with a corresponding donor strand in the manner described above, resulting in ligation of both donor strands which are reverse complementary with each other in the donor strand regions.

[0059] A ligase that is fused or recruited to an endonuclease, or supplied *in trans*, can utilize DNA as a splint, and a donor strand acts as the splint for the genomic flap generated by the endonuclease and another donor strand. In some aspects, the donor strand comprises: 5' donor strand - flap binding site - guide binding site (optional) 3'. The flap binding site on one donor strand (Donor2) can be reverse complementary to the genomic flap, while the optional guide binding site on Donor2 is reverse complementary to the optional donor binding site of a guide nucleic acid (Guide1), and the donor strand can be at least partially reverse complementary to a different donor strand (Donor1). The 5' end of this Donor1 and the 3' end of the genomic flap can be ligated using the flap binding site and donor strand of the Donor2 as a splint. Such 2-sided approach utilizing dual guide nucleic acids with different spacer sequences can be adopted with Donor2, which provides the splint at the first genomic site and can be ligated on its 5' end to a 3' end of a different genomic flap at a nick created using a second Replacer2 guide nucleic acid (Guide2) with a spacer sequence that targets a second site. The donor binding site on the second guide nucleic acid system can optionally recruit Donor1 via hybridization with its optional guide binding site, and the Donor1 acts as the DNA splint for ligation of Donor2 to the 3' end of the genomic flap at the target site of the second guide nucleic acid.

[0060] Following ligation, the remaining flaps of native genomic DNA can be excised via exogenously delivered or endogenous flap endonucleases or exonucleases. Examples of exogenous nucleases that can be introduced into the cell include human flap endonuclease 1 (hFEN1), human exonuclease 5 (hEXO5), T5 exonuclease, T7 exonuclease, exonuclease VIII, the flap endonuclease domain of *E. coli* PolI, RecJF, Lambda exonuclease, Xni (ExoIXI) from

Escherichia coli, SaFEN (*Staphylococcus aureus* FEN), nuclease BAL-31, or fragments thereof. The endonucleases or exonucleases can optionally be fused, recruited, or unfused to the RNA-guided endonuclease or DNA ligase by utilizing peptide linkers, heterodimerization domains, or two separate peptides, respectively.

[0061] In some aspects, the system, composition, or method described herein utilizes additional protein that binds to the cleaved or nicked site. For example, the system, composition, or method described herein can include Ku protein or Gam protein from bacteriophage Mu, where the binding of the Ku protein or Gam protein can increase ligation efficiency of the integration nucleic acid at the cleaved or nicked site.

[0062] A system or method described herein may use a nicking endonuclease and, therefore, does not generate double stranded breaks. Furthermore, the system described herein addresses the issue of poor editing efficiencies in nondividing cells through a mechanism of action which only depends on the exogenous components delivered to the cells using mRNA, viral vectors, guide nucleic acids, DNA, or peptides, or any other modalities. Therefore, the system does not require the presence of cell cycle-dependent endogenous cell processes or components such as HDR or dNTPS. As such, the system described herein allows efficiency that is not hindered in nondividing cells. Furthermore, the system enables replacement of both strands of a targeted region of the genome, which can increase editing efficiency.

[0063] A donor strand may contain a high degree of homology with the replaced genomic DNA. These donors may contain mutations to the genomic DNA such as pathogenic mutation correction, disabling of CRISPR protospacer adjacent motif (PAM) sites, disruption of the guide's spacer sequences, other substitution mutations, or a combination thereof. Additional substitution mutations may be included to increase donor-donor homology versus donor-genome homology to promote hybridization of donor strands and incorporation into the genome. Donor strands may also encode deletions or insertions of nucleotides, or may encode a complex combination of the above which then replaces the target genomic DNA. Optionally, guide and donor strands may be chemically modified using nucleic acid chemistries such as phosphorothioate bonds or 2'-O-methylation. Optionally, guide nucleic acids may include hairpin sequences. Optionally, any combination of guide nucleic acids, donor strands, and proteins can be complexed, using an annealing reaction (gradual reduction in temperature) for example, prior to delivering the editing components to the cell.

[0064] Protein components (e.g. nicking Cas9, ligase) may be modified using nuclear localization signals, cell penetrating peptides, or chromatin disrupting peptides in order to improve delivery efficiency to genomic targets.

[0065] The predominant cellular DNA repair pathway for resolving small (< 13nt) mismatches between genomic DNA strands is mismatch repair (MMR). For single stranded donor ligation, the ligated donor strand forms a DNA heteroduplex with the reverse complementary genomic DNA strand. This may also occur with competitive hybridization between ligated donor strand strands and genomic DNA strands. In these cases, MMR activity can excise and revert mismatches in the donor strand using the genomic strand as a template, resulting in reduced editing. Expression of dominant negative versions of MMR proteins has been shown to inhibit the MMR pathway and improve editing outcome in cases where similar DNA heteroduplexes are generated. In some aspects, dominant negative MMR peptides such as MSH2 (G674A) and MLH1 (del754-756) may be delivered as part of the system described herein to improve genomic editing capability, particularly in cells which overexpress the MMR pathway. In some aspects, these dominant negative MMR peptides can be delivered as a fusion (e.g., fused with any component of the system described herein), recruited, or as separate peptides.

Endonucleases

[0066] Disclosed herein are endonucleases. The endonuclease may be included in a composition, system or method disclosed herein. The endonuclease may be recombinant. The endonuclease may be coupled to a ligase. The endonuclease may be coupled directly or indirectly to the ligase. The coupling may be covalent or non-covalent. The endonuclease may be bound or connected to a ligase. The endonuclease may be recruited to, be part of a fusion protein with, or be used in conjunction with the ligase. The endonuclease may be heterologous. Heterologous may indicate a source from without a cell. Where a heterologous endonuclease is described, a non-heterologous (e.g. endogenous) endonuclease may be used in some instances. The endonuclease may be encoded in a cell. The endonuclease may be delivered to the cell in trans. The endonuclease may catalyze cleavage of a phosphate bond within an integrating nucleic acid. The endonuclease may be guided by a guide nucleic acid to cleave or nick a target nucleic acid for ligation of an integrating nucleic acid at the cleavage or nick site. The endonuclease may include any aspect included in **Fig. 1A-6C**.

[0067] The endonuclease may be non-naturally occurring. The endonuclease may be engineered. The endonuclease may be synthetic. The endonuclease may be pre-synthesized. The endonuclease may be added to a subject or a cell. The endonuclease may be encoded by a nucleic acid. The encoding nucleic acid may be engineered, synthetic, or added to a subject or a cell.

[0068] At least part of the endonuclease may be included in a first polypeptide. At least part of the endonuclease may be included in a second polypeptide. The endonuclease may be split into two or more polypeptides bound together. The first polypeptide may include an N-terminal

portion of the endonuclease. The first polypeptide may include a C-terminal portion of the endonuclease. The second polypeptide may include the N-terminal portion of the endonuclease. The second polypeptide may include the C-terminal portion of the endonuclease. The first or second polypeptide comprising a part of the endonuclease may be fused with at least part, or the whole, of the ligase.

[0069] Described herein, in some aspects, is a system comprising at least one endonuclease. In some aspects, the endonuclease is a programmable endonuclease, where the endonuclease can be complexed with and directed by a guide nucleic acid described herein to a genomic locus. The endonuclease may bind DNA. In some aspects, the endonuclease is a RNA-guided endonuclease. In some aspects, the endonuclease can introduce a single-stranded break. Examples of RNA-guided endonucleases can include CRISPR/Cas endonucleases (e.g., class 2 CRISPR/Cas endonucleases such as a type II, type V, or type VI CRISPR/Cas endonucleases). A CRISPR/Cas endonuclease is also referred to as a CRISPR/Cas effector polypeptide. A suitable endonuclease is a CRISPR/Cas endonuclease (e.g., a class 2 CRISPR/Cas endonuclease such as a type II, type V, or type VI CRISPR/Cas endonuclease). In some cases, a suitable RNA-guided endonuclease is a class 2 CRISPR/Cas endonuclease. In some cases, a suitable RNA-guided endonuclease is a class 2 type II CRISPR/Cas endonuclease (e.g., a Cas9 protein). In some cases, an endonuclease includes a class 2 type V CRISPR/Cas endonuclease (e.g., a Cpf1 protein, a C2c1 protein, or a C2c3 protein). In some cases, a suitable RNA-guided endonuclease is a class 2 type VI CRISPR/Cas endonuclease (e.g., a C2c2 protein; also referred to as a “Cas13a” protein). Also suitable for use is a CasX protein. Also suitable for use is a CasY protein. In some aspects, the endonuclease can include any one of the Cas described herein complexed with a guide nucleic acid (e.g., a gRNA) as an RNP complex.

[0070] In some cases, the endonuclease is a Type II CRISPR/Cas endonuclease. In some cases, the endonuclease is a Cas9. Cas9 functions as an RNA-guided endonuclease that uses a dual-guide RNA having a crRNA and trans-activating crRNA (tracrRNA) for target recognition and cleavage by a mechanism involving two nuclease active sites in Cas9 that together generate double-stranded DNA breaks (DSBs), or can individually generate single-stranded DNA breaks (SSBs). The Type II CRISPR endonuclease Cas9 and engineered dual- (dgRNA) or single guide RNA (sgRNA) form a ribonucleoprotein (RNP) complex that can be targeted to a desired DNA sequence. Guided by a dual-RNA complex or a chimeric single-guide RNA, Cas9 generates site-specific DSBs or SSBs within double-stranded DNA (dsDNA) target nucleic acids, which are repaired either by non-homologous end joining (NHEJ) or homology-directed recombination (HDR). The Cas9 can be guided to a target site (e.g., stabilized at a target site) within a target nucleic acid sequence by virtue of its association with the RNA-binding segment

of the Cas9 to guide RNA. A Cas9 protein can bind and/or modify (e.g., cleave, nick, methylate, demethylate, etc.) a target nucleic acid and/or a polypeptide associated with target nucleic acid (e.g., methylation or acetylation of a histone tail)(e.g., when the Cas9 protein includes a fusion partner with an activity). In some cases, the Cas9 protein is a naturally-occurring protein (e.g., naturally occurs in bacterial and/or archaeal cells). In other cases, the Cas9 protein is not a naturally-occurring polypeptide (e.g., the Cas9 protein is a variant Cas9 protein, a chimeric protein, and the like).

[0071] Naturally occurring Cas9 proteins may bind a Cas9 guide RNA, are thereby directed to a specific sequence within a target nucleic acid (a target site), and cleave the target nucleic acid (e.g., cleave dsDNA to generate a double strand break, cleave ssDNA, cleave ssRNA, etc.). A chimeric Cas9 protein may include a fusion protein comprising a Cas9 polypeptide fused to a heterologous protein (referred to as a fusion partner), where the heterologous protein provides an activity (e.g., one that is not provided by the Cas9 protein). The fusion partner can provide an activity, e.g., enzymatic activity (e.g., nuclease activity, activity for DNA and/or RNA methylation, activity for DNA and/or RNA cleavage, activity for histone acetylation, activity for histone methylation, activity for RNA modification, activity for RNA-binding, activity for RNA splicing etc.). In some cases, a portion of the Cas9 protein (e.g., the RuvC domain and/or the HNH domain) exhibits reduced nuclease activity relative to the corresponding portion of a wild type Cas9 protein (e.g., in some cases the Cas9 protein is a nickase). In some cases, the Cas9 protein is enzymatically inactive, or has reduced enzymatic activity relative to a wild-type Cas9 protein (e.g., relative to *Streptococcus pyogenes* Cas9). In some cases, the Cas9 is a Cas9 nickase. The Cas9 nickase can be generated by mutating a Cas9 nuclease domain. Non-limiting example of the Cas9 nickase can include SpCas9, SaCas9, CjCas9, GeoCas9, HpaCas9, and NmeCas9. In some aspects, the endonuclease described herein comprises any one of the Cas9 in **Table 1**. In some aspects, the endonuclease described herein comprises a polypeptide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of the Cas9 in **Table 1**.

Table 1. Non-limiting examples of Cas9 polypeptide sequence

Name	Cas9 polypeptide sequence	SEQ ID NO:
SpyCas9	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK NGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA QIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF LDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIH DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV DELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKM NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDV RKMIASEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN GETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKK LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIKLPKYSLF ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNK HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSIITGLYETRIDLSQLGGD	1
Nicking SpyCas9 (H840A)	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFHRLLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK NGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA QIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF LDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIH DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV DELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD	2

Name	Cas9 polypeptide sequence	SEQ ID NO:
	YDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDV RKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN GETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKK LKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENKRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNK HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSIITGLYETRIDLSQLGGD	
Nicking SpyCas9 (H840A) R221K N394K	MDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK KLVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRKLENLIAQLPGEKK NGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLKREDDL RKQRTFDNGSIPHQIHLGELH AILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDF LDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIH DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV DELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSD YDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQ ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDV RKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN GETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKK LKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENKRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNK HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSIITGLYETRIDLSQLGGD	3
Nicking SpyCas9 (D10A)	MDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK KLVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK NGLFGNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY	4

Name	Cas9 polypeptide sequence	SEQ ID NO:
	KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLL YEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIH DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKE LGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMMK NYWRQLLNAKLITQRKFDNLTKAERGGSELDAKAGFIKRLVETRO ITKHVAQILDSRMNTKYDENDKLIREVKVITLKSLLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDV RKMIASEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN GETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILP KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKK LKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGP EDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNK HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD ATLIHQSIITGLYETRIDLSQLGGD	
SauCas9	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGR RSKRGARRLKRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQIS RNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWY EMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIENVFQKQKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF TNLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSEDIQEELTNLN SELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYG LPNDIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENA KYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSF DNSFNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLA KGKGRISKTKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMN LLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAE DALIIANADFIFKEWKLDKAKKVMENQMFEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKNRELINDTLYSTRKDDK GNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLK LIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNK LNAHLDITDDYPNSRNKVVKLSLKPYPYFDVYLDNGVYKFVTVKNL DVIKKENYYEVNSKCYEEAKKLKISNQAEFIASFYNNDLIKINGEL YRVIGVNNDDLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQS IKKYSTDILGNLYEVKSKKHPQIIKKG*	5

Name	Cas9 polypeptide sequence	SEQ ID NO:
Nicking SauCas9 (N580A)	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGR RSKRGARRLKRRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEARV KGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQIS RNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWY EMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIENVFQKQKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF TNLKVYHDIKDITARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLN SELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYG LPNDIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENA KYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEVVDHIIPRSVSF DNSFNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMN LLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAE DALIIANADFIFKEWKLDKAKKVMENQMFEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKNRELINDTLYSTRKDDK GNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPPQTYQKLK LIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNK LNAHLDITDDYPNSRNKVVKLKPYRFDVYLDNGVYKFVTVKNL DVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYNNDLIKINGEL YRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQS IKKYSTDILGNLYEVKSKKHPQIIKK*	6
KKH- SaCas9	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEG RRSKRGARRLKRRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEAR VKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQI SRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLL KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEKSPFGWKDIKEWY EMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIENVFQKQKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF TNLKVYHDIKDITARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLN SELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYG LPNDIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENA KYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEVVDHIIPRSVSF DNSFNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMN LLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAE DALIIANADFIFKEWKLDKAKKVMENQMFEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDK GNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPPQTYQKLK LIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNK LNAHLDITDDYPNSRNKVVKLKPYRFDVYLDNGVYKFVTVKNL DVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYKNDLIKINGEL YRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPHIKTIASKTQS IKKYSTDILGNLYEVKSKKHPQIIKK	7
Nicking KKH- SaCas9 (N580A)	MGKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEG RRSKRGARRLKRRRRRHRIQRVKLLFDYNLLTDHSELSGINPYEAR VKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQI SRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLL	8

Name	Cas9 polypeptide sequence	SEQ ID NO:
	KVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWY EMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKL EYYEKFQIIEENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEF TNLKVYHDIKDITARKEIIEAELLDDQIAKILTIYQSSEDIQEELTNLN SELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNR LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYG LPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIIRTGKENA KYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSF DNSFNKVLVKQEEASKKGNRTPFQYLSSSDSKISYETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMN LLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAE DALIIANADFIFKEWKKLDKAKKVMENQMFEKQAESMPEIETEQE YKEIFITPHQIKHIKDFKDYKYSHRVDKKNRKLINDTLYSTRKDDK GNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHDPTQYQKLK LIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKIKYYGNK LNAHLDITDDYPNSRNKVVKLSLKPYPYFDVYLDNGVYKFVTVKNL DVIKKENYYEVNSKCYEEAKLKKISNQAEFIASFYKNDLIKINGEL YRVIGVNNDDLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQS IKKYSTDILGNLYEVKSKKHPQIIKK	
CjCas9	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPR RLARSARKRLARRKARLNHLKHLIANEFKLNIEDYQSFDESLAKA YKGSLSIPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDK EKGAILKAIQNEEKLANYQSVGEYLYKEYFQKFENSKEFTNVRN KKEYSERCIAQSFLKDELKLIFKKQREFGFSFSSKFEEVLSVAFYK RALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKN TEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEK GTYFIEFKKYKEFIKALGEHNSQDDLNEIAKDITLIKDEIKLKKALA KYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACN ELNLKVAINEDKKDFLPAFNETYYKDEVTPNPVVLRAIKEYRKVLNA LLKKYGVKHINIELAREVGKNHSQRAKIEKEQNENYKAKKDAEL ECEKGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEID HIYPYSRSFDDSYMNVLVFTKQKQEKLNQTPFEAFGNDSAKWQK IEVLAKNLPKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLN YTKDYLDLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTW GFSKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYA KKISELDYKNKRKFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALH EETFRKEEFYQSYGGKEGVKALELGKIRK VNGKIVKNGDMFRV DIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILM DENYEFCSLYKDSLILIQTKDMQEPEFVYNAFTSSTVSLIVSKHD NKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVT KAEFRQREDFKK	9

Name	Cas9 polypeptide sequence	SEQ ID NO:
GeoCas 9	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLA LPRRLARSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEKHEI DVWQLRVEALDRKLNDELARVLLHLAKRRGFKSNRKSERSNKEN STMLKHIEENRAILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIAR DDLEREIRLIFSKQREFGNMSCTEEFENEYITIWASQRPVASKDDIEK KVGFCFTFEPKEKRAPKATYTFQSFIAWEHINKLRLISPSGARGLTDE ERRLLYEAFQKNKITYHDIRTLLHLPDDTYFKGIVYDRGESRKQN ENIRFLELDAYHQIRKAVDKVYGGKSSSFLPIDFTFGYALTLFKD DADIHSYLRNEYEQNGKRMPNLANKVVDNELIEELLNLSFTKFGHL SLKALRSILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPNIPPI ANPVVMRALTQARKVVNAIIKKYGGSPVSIHIELARDLSQTFDERRKT KKEQDENRKKNETAIRQLMEYGLTLNPTGHDIVKFKLWSEQNGRC AYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNKVLVLTRENRE KGNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLRLHYDE NEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGR VTAHLRSRWEFNKNREESDLHHAVIDAVIVACTTPSDIAKVTAIFYQR REQNKELAKKTEPHFPQPWPHFADELARLSKHPKESIKALNLGNY DDQKLESLLQPVFVSRMPKRSVTGAHQETLRRYVGIDERSGKIQT VKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQRILLEHNNDPKKAF QEPLYKPKKNGEPGPVIRTVKIIDTKNQVIPLNDGKTVAYNSNIVRV DVFEKDGKYYCVPVYTMIMKILPNKAIEPNKPYSEWKEMTEDY TFRFSLYPNDLIRIELPREKTVKTAAGEEINVKDVVFVYKTIIDSANG GLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNIYKVRGEKRVGL ASSAHSKPGKTIRPLQSTRD	10
HpaCas 9	MENKNLNYILGLDLGIASVGWAVVEIDEKENPLRLIDVGVRTFERA EVPKTGESLALSRRRLARSARRLTQRRVARLKKAKRLLKSENILLSTD ERLPHQVWQLRVEGLDHLKLERQEWAAVLLHLIKHRGYLSQRKNES KSENKELGALLSGVDNNHKLLQQATYRSPAELAVKKFEVEEGHIRN QQGAYTHTFSRLDLLAEMELLFSRQQHFGNPFASEKLENLTALLM WQKPALSGEAILKMLGKCTFEDEYKAAKNTYSAERFVWITKLNLL RIQENGLERALNDNERLALMEQPYDKNRLFYSQVRSILKLSDEAIFK GLRYSGEDKKAIEKAVLMEMKAYHQIRKVLEGNLKAWEAELK ANPTLLDEIGTAFSLYKTDEDISAYLAGKLSQPVLNALLNLSFDKFI QLSLKALYKLLPLMQQGLRYDEACREIYGDHYGKKTEENHHFLPQI PADEIRNPVVLRLTLTQARKVINGVVRLYGGSPARIHETGREVGKSYK DRRELEKRQEENRKQRENAIKEFKEYFPHFAGEPKAKDILKMRLYK QQNAKCLYSGKPIELHRLLEKGYVEVDHALPFSRTWDDSFNNKVL VLANENQNKGNLTPFEWLDGKHNSERWRAFKALVETSAFPYAKK QRILSQKLDEKGFIERNLNDTRYVARFLCNFIADNMHLTGEGKRKV FASNGQITALLRSRWGLAKSREDNDRHHALDAVVVACSTVAMQQ KITRFVRFEAGDVFTGERIDRETGEIPLHFPTPWQFFKQEVEIRIFSD NPKLELENRLPDRPQANHEFVQPLFVSRMPTRKMTGQGHMETVKS AKRLNEGISVIKMPLTKLKLKDLELMVNREREKDLYDTLKARLEAF NDDPAKAFAPFIKKGGAIVKSVRVEQIQKSGVLVREGNGVADNAS MVRVDVFTKGGKYFLVPIYTQVAKGILPNKAATQYKDEEDWEV MDNSATFKFSLHPNDLVKLVTKKKTILGYFNGLNRATGNIDIKEHD LDKSKGKGQIFEGVGIKLALSFEKYQVDELGKNIRLCKPSKRQPV	11

Name	Cas9 polypeptide sequence	SEQ ID NO:
SmuCas 9	MMMEKFHYVLGLDLGIA SVGWAAIEIDKETETSIGLLDCGVRTFER AEVPKTGDSLAKARREARSTRRLIRRRSHRLLRLKRLKREIFRQPE TFKDLPINAWQLRVKGLDSRLNEYEWAAVLLHLVKHRGYLSQRKS EMSETDSKSEMGRLLAGVAENHQLLQQEQYRTPAELALKKFVKHF RNKGGDYAHTFNRLDLQAEHLHLLFQKQRELGNPFTSPELERQVDDL LMTQRSALQGDAILKMLGHCGFEPEQFKAANKNTFSAERFIWLTCLN NLRIQDQGKERALTADERTKLLDEPYKSKSLTYAQVRKLLSLPQTA IFKGLRYDLEHDKKAENSTLMEMKSYHNIRQTLEKSGLKTEWQSA TQPEILDAIGTAFSIYKTDEDISHELKTCRLPENVLNELLKNINFDGFI QLSLTALRKILPLMEQGYRYDEACTQIYGNHHSGLQQESKQFLPHI PIDDVNRNPVVFRTLTQARKVVNAIIRRYGSPARVHIEMARELGKSKS DRDRIEKQQKNKKERENAVAKFKEDFPDFVGEPRGKDILKMRLY EQQHKGKCLYSGHDIDINRLNEKGYVEIDHALPFSRTWDDSQNNKVL VLGSENQNKRNQTPDEYLDGANNSSQRWLEFQARVQTCHFSYGKK QRIQLAKLDDDETEKGFLERNLNDTRYIARFMCQFVQENLYLTGKGG RLVFASNGGMTATLRNLWGLRKVREDNDRHHALDAIVVACSTAS MQQKITKAFQRHESIEYVDTETGEVKFRIPQPWDFFRQEV MIRVFS QPCEDLVEKLSARPEALHDNVTPLFVSRAPNRKMSGQGHLETIKSA KRLSEENSMVKKPLTTLKLDIPEIVGYPSREPQLYAALKTRLETHD DDPIKAFAPFYKPNKNGELGALVRSVRVKG VQNTGVMVHDGKGI ADNATMVRVDVYTKAGKNYLVPVYVWQVAQGILPNRAVTSGKSE ADWDLIDESFEFKFSLSRGDLVEMISNKGRIFGYNGLDRANGSIGI REHDLEKSKGKDGVRVGVKTATAFNKYHVDPLGKEIHRCSSSEPR PTLKI KSKK	12
NmeCas 9	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGV RVFER AEVPKTGDSLAMARRLARSVRRLTRRRAHRLRLTRRLKREGVLQ AANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRG YLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNK FEKESGHIRNQRSDYSHTFSRKDLQAEILLFEKQKEFGNPHVSGGL KEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERF IWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKSLTYAQARKL LGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDK KSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISF DKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIY LPPIPADEIRNPVVLRAALSQARKVINGVVRRYGSPARIHIETAREVGK SFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKRL YEQQHKGKCLYSGKEINLGRLEKGYVEIDHALPFSRTWDDSFNNK VLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETS RFPRSK KQRILLQKFDEDFGKERNLNDTRYVNRFLCQFVADRMRLTGKGKK RVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAM QQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEV MIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFV SRAPNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLKDEK MVNREREPKLYEALKARLEAHKDDPAKAFAPFYKYDKAGNRTQ QVKAVRVEQVQKTGVWVRNHNHGIADNATMVRVDVFEKGDKYYL VPIYSWQVAKGILPDRAVVQKDEEDWQLIDDSFNFKFSLHPNDLV EVITKKARMFGYFASCHRG TGNINIRIHDLDHKIGKNGILEGIGVKT ALSFQKYQIDELGKEIRPCRLKKRPPVR	13

[0072] Some aspects include an endonuclease such as an RNA-guided endonuclease. The RNA-guided endonuclease may comprise a class II CRISPR/Cas endonuclease. The RNA-guided endonuclease may comprise a Cas9 endonuclease. The RNA-guided endonuclease may comprise a nickase. The RNA-guided endonuclease may comprise an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 1-13, or a functional fragment thereof.

[0073] The endonuclease may introduce a single-strand break in a target nucleic acid. The endonuclease may introduce a single-strand break in a target nucleic acid without cleaving a strand opposite the single strand break. The endonuclease may include a nickase. In some instances, the endonuclease may exclude an endonuclease that introduces a double strand break. The endonuclease may exclude a restriction enzyme.

[0074] The endonuclease may be included as part of a fusion protein. In some cases, an endonuclease is a fusion protein that is fused to a heterologous polypeptide such as the heterologous ligase described herein. The heterologous polypeptide may include a fusion partner. The fusion protein may include a fusion partner such as a DNA ligase, a nuclear localization signal, chromatin modifying domain, cell penetrating peptide, or tag polypeptide. The fusion protein may include one or more fusion partner. The fusion protein may include a ligase. The fusion protein may include a nuclear localization signal, chromatin modifying domain, cell penetrating peptide, or tag polypeptide.

[0075] The fusion partner may be connected to the N-terminus of the endonuclease. The fusion partner may be connected to the C-terminus of the endonuclease. The endonuclease may be connected at an N-terminus or a C-terminus to a linker. The fusion partner may be connected by the fusion partner's N-terminus or C-terminus. The fusion partner may be connected by the fusion partner's N-terminus to the endonuclease. The fusion partner may be connected by the fusion partner's C-terminus to the endonuclease. The fusion partner may be connected at an N-terminus or a C-terminus to a linker.

[0076] In some cases, the endonuclease comprises a linker, where the linker covalently connects the endonuclease to the heterologous polypeptide. The linker may connect the endonuclease to any fusion partner. A linker may also connect any fusion partner to another fusion partner. The linker polypeptide may have any of a variety of amino acid sequences. Proteins can be joined by a spacer peptide, generally of a flexible nature, although other chemical linkages are not excluded. Suitable linkers include polypeptides of between 4 amino acids and 40 amino acids in length, or between 4 amino acids and 25 amino acids in length. These linkers can be produced by using synthetic, linker-encoding oligonucleotides to couple the proteins, or can be encoded by a nucleic acid sequence encoding the fusion protein. Peptide

linkers with a degree of flexibility can be used. The linking peptides may have virtually any amino acid sequence, bearing in mind that the preferred linkers will have a sequence that results in a generally flexible peptide. The use of small amino acids, such as glycine and alanine, are of use in creating a flexible peptide. The creation of such sequences is routine to those of skill in the art. A variety of different linkers are commercially available and are considered suitable for use. Examples of linker polypeptides include glycine polymers (G)_n, glycine-serine polymers (including, for example, (GS)_n, (GSGGS)_n, (GGSGGS)_n, and (GGGS)_n, where n is an integer of at least one); glycine-alanine polymers; and alanine-serine polymers. Exemplary linkers can comprise amino acid sequences including, but not limited to, GGSG, GGSGG, GSGSG, GSGGG, GGGSG, GSSSG, and the like. Also suitable is a linker having the sequence (GGGGS)_n, where n is an integer of from 1 to 10 (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10). The ordinarily skilled artisan will recognize that design of a peptide conjugated to any desired element can include linkers that are all or partially flexible, such that the linker can include a flexible linker as well as one or more portions that confer less flexible structure.

[0077] One or more linkers may be included in a fusion protein. 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 linkers, or a range of linkers defined by any two of the aforementioned integers, may be included in the fusion protein. A linker may connect to an N-terminal end of at least part of the endonuclease. A linker may connect to an N-terminal end of at least part of a fusion partner. A linker may connect to an N-terminal end of at least part of a fusion ligase. A linker may connect to an N-terminal end of a nuclear localization signal. A linker may connect to an N-terminal end of a chromatin modifying domain. A linker may connect to an N-terminal end of a cell penetrating peptide. A linker may connect to an N-terminal end of a tag polypeptide. A linker may connect to a C-terminal end of at least part of the endonuclease. A linker may connect to a C-terminal end of at least part of a fusion partner. A linker may connect to a C-terminal end of at least part of a fusion ligase. A linker may connect to a C-terminal end of a nuclear localization signal. A linker may connect to a C-terminal end of a chromatin modifying domain. A linker may connect to a C-terminal end of a cell penetrating peptide. A linker may connect to a C-terminal end of a tag polypeptide.

[0078] A linker may comprise a number or range of amino acids or residues. The linker may include at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 13, at least 14, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 amino acid residues. The linker may, in some aspects, include no more than 1, no more than 2, no more than 3, no more than 4, no more than 5, no more than 6, no more than 7, no more than 8, no more than 9, no more than

10, no more than 12, no more than 13, no more than 14, no more than 15, no more than 20, no more than 25, no more than 30, no more than 35, no more than 40, no more than 45, no more than 50, no more than 55, no more than 60, no more than 65, no more than 70, no more than 75, no more than 80, no more than 85, no more than 90, no more than 95, or no more than 100 amino acid residues. A linker may include 1-10 amino acids, 1-25 amino acids, or 1-100 amino acids.

[0079] Linkers may be included anywhere in a polypeptide chain or protein described herein. For example, a linker may separate an endonuclease from a ligase. A linker may separate an endonuclease from a nuclear localization signal, a chromatin modifying domain, a cell penetrating peptide, or a tag polypeptide.

[0080] In some cases, the endonuclease comprises a nuclear localization sequence (e.g., one or more nuclear localization signals or NLSs for targeting to the nucleus). In some aspects, the NLS described herein comprises any one of the NLS in **Table 2**. In some aspects, the NLS described herein comprises a polypeptide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of NLS in **Table 2**.

Table 2. Non-limiting examples of NLS polypeptide sequence

Name	NLS polypeptide sequence	SEQ ID NO:
NLS1	KRTADGSEFESPKKKRKV	14
NLS2	SGGSKRTADSQHSTPPKTKRKVEFEPKKKKRKV	15
NLS3	KRPAATKKAGQAKKKK	16
NLS4	KKTELQTTNAENKTKKL	17
NLS5	KRGINDRNFWRGENGRKTR	18
NLS6	RKSGKIAAIVVKRPRK	19
NLS7	PKKKRKV	20
NLS8	MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	21
SGGSx2-bpNLS-SGGSx2	SGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGS	22
SGGSx2-XTEN16-SGGSx2	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	23
SGGSx10	SGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSS	24

[0081] A polynucleotide encoding an NLS polypeptide may be used. An example of such a polynucleotide may be SGGs2-bpNLS-SGGs2:

TCCGGCGGAAGCTCTGGTGGCAGCAAGCGGACCGCCGACGGCTCTGAATTCGAGA
GCCCTAAGAAGAAAAGAAAGGTGAGCGGAGGCTCTAGCGGCGGAAGC (SEQ ID
NO: 25).

[0082] In some aspects, the endonuclease comprises a dimerization domain. The dimerization domain can be located at the N-terminus or C-terminus of the endonuclease. In some aspects, the dimerization domain allows the endonuclease to form a heterodimer with another polypeptide (e.g., the heterologous ligase). In some aspects, the dimerization domain allows the endonuclease to be functionally coupled with another polypeptide. Non-limiting examples of the dimerization domains can include a leucine zipper, an FKBP, an FRB, a Calcineurin A, a CyP-Fas, a GyrB, a GAI, a GID1, a SNAP tag, a Halo tag, a Bcl-xL, a Fab, a LOV domain, or SpyTag/SpyCatcher. Other example of dimerization domain can include an antibody such as anyone of heavy chain domain 2 (CH2) of IgM (MHD2) or IgE (EHD2), immunoglobulin Fc region, heavy chain domain 3 (CH3) of IgG or IgA, heavy chain domain 4 (CH4) of IgM or IgE, Fab, Fab2, leucine zipper motifs, barnase-barstar dimers, miniantibodies, or ZIP miniantibodies. In some aspects, the dimerization domain described herein comprises any one of the dimerization domain in **Table 3**. In some aspects, the dimerization domain described herein comprises a polypeptide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of dimerization domain in **Table 3**.

Table 3. Non-limiting examples of dimerization domain sequence

Name	Dimerization domain sequence	SEQ ID NO:
Leucine zipper EE12RR345L	LEIEAAFLERENTALETRVAELRQVRQLNRVVSQYRTRYGPL GGGK	26
Leucine zipper RR12EE345L	LEIRAAFLRQRNTALRTEVAELEQEVQRLENEVSQYETRYGPL GGGK	27

[0083] In some aspects, the endonuclease comprises at least one additional domain. In some aspects, the at least one additional domain is a functional domain. For example, the functional domain can comprises a chromatin modifying domain or a cell penetrating peptide. In some aspects, the chromatin modifying domain described herein comprises any one of the chromatin modifying domain in **Table 4**. In some aspects, the chromatin modifying domain described herein comprises a polypeptide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of chromatin modifying domain in **Table 4**.

Table 4. Non-limiting examples of chromatin modifying domain polypeptide sequence

Name	Chromatin modifying domain polypeptide sequence	SEQ ID NO:
H1G (histone H1 central globular domain)	STDHPKYSMDIVAAIQAENRAGSSRSIQKYIKSHYKVGENA DSQIKLSIKRLVTTGVLKQTKGVGASGSFRLAKSDEP	28

Name	Chromatin modifying domain polypeptide sequence	SEQ ID NO:
HMGB1	MGKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFS KKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKG ETKKKFKDPNAPKRPPSAFFLFCSEYRPIKGEHPGLSIGDVAK KLGEMWNNTAADDKQPYEKKA AKLKEKYEKDIAAYRAKGKP DAAKKGVVKA EKSKKKKEEEEEDEEDEDEDEDEDEDEDEDE DDDE	29
HMGB2	MGKGDPNKPRGKMSSYAFFVQTCREEHKKKHPDSSVNFSEFS KKCSERWKTMSAKEKSKFEDMAKSDKARYDREMKNYVPPKG DKKGKKKDPNAPKRPPSAFFLFCSEHRPIKSEHPGLSIGDTAK KLGEMWSEQSAKDKQPYEQKA AKLKEKYEKDIAAYRAKGKS EAGKKGPGRPTGSKKKNEPEDEEEEEDEDEDEDEDEDEDEE	30
HMGB3	MAKGDPKKPKGKMSAYAFFVQTCREEHKKKNPEVPVNFAEFS KKCSERWKTMSGKEKSKFEDMAKADKVRDREMKDYGPAK GGKKKKDPNAPKRPPSGFFLFCSEFRPIKSTNPGISIGDVAKKL GEMWNNLNDSEKQPYITKA AKLKEKYEKDVADYKSKGKFDG AKGPAKVARKKVEEEDEEEEEEEEEEEEEEEEDE	31
HMGN1 (HN1)	MPKRKVSSAEGAAKEPKRRSARLSAKPPAKVEAKPKKAAAK DKSSDKKVQTKGKRGAKGKQAEVANQETKEDLPAENGETKTE ESPASDEAGEKEAKSD	32
HMGN2	MPKRKAEGDAKGDKAKVKDEPQRRSARLSAKPAPPKPEPKPK KAPAKKGEKVPKGKKGKADAGKEGNNPAENGDAKTDQAQKA EGAGDAK	33

[0084] In some aspects, the cell penetrating peptide described herein comprises any one of the cell penetrating peptide in **Table 5**. In some aspects, the cell penetrating peptide described herein comprises a polypeptide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of cell penetrating peptide in **Table 5**.

Table 5. Non-limiting examples of cell penetrating peptide polypeptide sequence

Name	Cell penetrating peptide sequence	SEQ ID NO:
Penetratin	RQIKIYFQNRRMKWKK	34
TAT	RKKRRQRRR	35
R8	RRRRRRRR	36
DPV3	RKKRRRESRKKRRRES	37
DPV6	GRPRESGKKRKRRLKP	38
R9-TAT	GRRRRRRRRRPPQ	39
pVEC	LLIILRRRIRKQAHAAHSK	40
ARF(19-31)	RVRVFVVWHIPRLT	41
MPG	GALFLGFLGAAGSTMGAWSQPKKRKV	42
Transportan	GWTLSAGYLLGKINLKALAALAKKIL	43
Bip4	VSALK	44
C105Y	CSIPPEVKFNPVYLI	45
Melittin	GIGAVLKVLTTGLPALISWIKRKRQQ	46
gH625	HGLASTLTRWAHYNALIRAF	47

[0085] In some aspects, the endonuclease comprises a tag, where the tag can be used for increasing expression, identifying, or purifying the endonuclease. In some aspects, the tag described herein comprises any one of the tag sequence in **Table 6**. In some aspects, the tag described herein comprises a polypeptide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of the tag sequence in **Table 6**.

Table 6. Non-limiting examples of tag polypeptide sequence

Name	Tag polypeptide sequence	SEQ ID NO:
FLAG	DYKDDDDK	48
His-Tag	HHHHHH	49
CBP	KRRWKKNFIAVSAANRFKKISSSGAL	50
MBP	MKIKTGARILALSALTMMFSASALAKIEEGKLVWINGDKGY NGLAEVGGKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDII FWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNG KLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKS ALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNA GAKAGLTFLVDLIKNNHMNADTDYSIAEAAFNKGETAMTING PWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASP NKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK DPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQ TVDEALKDAQTRITK	51
Myc	EQKLISEEDL	52
GST	MKLFYKPGACSLASHITLRESGKDFTLVSVDLMKKRLNGDN YFAVNPQGQVPALLDDGTLLTEGVAIMQYLADSVDRQLLA PVNSISRYKTIEWLNYIATELHKGFTPLFRPDTPEEYKSTVRAQ LEKKLQYVNEALKDEHWICGQRFITADAYLFTVLRWAYAVKL NLEGLEHIAAFMQRMAERPEVQDALSAEGLK	53
HA	YPYDVPDYA	54
HA	YAYDVPDYA	210
HA	YDVPDYASL	211

[0086] In some embodiments, the endonuclease can be expressed as split construct as one or more exteins fused to one or more inteins. Intein technology may be used to deliver large proteins into a cell by expressing the protein as two or more shorter peptide segments (exteins). Each extein may be expressed as a fusion with an intein peptide (e.g., an Npu C intein or an Npu N intein). An intein may autocatalyze fusion of two or more exteins and may autocatalyze excision of the intein from its corresponding extein. The result may be a protein complex comprising a first extein fused to a second extein and lacking inteins. An intein may be positioned N-terminal of the extein, or an intein may be positioned C-terminal of the extein. An extein may comprise a cysteine residue positioned adjacent to the intein (e.g., at the C-terminal end of an extein with an intein fused to the C-terminal end of the extein). The Cas nickase may be expressed as two or more segments. A first of the Cas nickase segment may comprise an N-

terminal portion of the Cas nickase. A first segment of the Cas nickase may comprise a first intein. A second segment of the Cas nickase may comprise a C-terminal portion of the Cas nickase. A second segment of the Cas nickase may comprise a second intein. An intein may be fused to a C-terminus of an N-terminal portion of the Cas nickase. An intein may be fused to an N-terminus of a C-terminal portion of the Cas nickase. A nucleic acid sequence encoding an extein-intein fusion may fit into a delivery vector (e.g., an adeno-associated virus (AAV) vector).

DNA Ligases

[0087] Disclosed herein are ligases. The ligase may be or include a DNA ligase. The ligase may be included in a composition, system or method disclosed herein. The ligase may be recombinant. The ligase may be coupled to the endonuclease. The ligase may be coupled directly or indirectly to the endonuclease. The coupling may be covalent or non-covalent. The ligase may be bound or connected to the endonuclease. The ligase may be recruited to, be part of a fusion protein with, or be used in conjunction with an endonuclease. The ligase may be heterologous. The ligase may be endogenous. Where a heterologous ligase is described, a non-heterologous (e.g. endogenous) ligase may be used in some cases. The ligase may be encoded in a cell. The ligase may be delivered to the cell in trans. The ligase may form a phosphodiester bond by joining two nucleic acid ends together. The ligase may join an end (e.g. 5' or 3' end) of a target nucleic acid to an integrating nucleic acid (e.g. a 3' or 5' end of the integrating nucleic acid). The ligase ligates an integrating nucleic acid to a cleaved or nicked end of a target nucleic acid where the cleaved or nicked end has been generated by an endonuclease such as an RNA-guided endonuclease. The ligase may include any aspect included in **Fig. 1A-6C**.

[0088] The ligase may be non-naturally occurring. The ligase may be engineered. The ligase may be synthetic. The ligase may be pre-synthesized. The ligase may be added to a subject or a cell. The ligase may be encoded by a nucleic acid. The encoding nucleic acid may be engineered, synthetic, or added to a subject or a cell.

[0089] At least part of the ligase may be included in a first polypeptide. At least part of the ligase may be included in a second polypeptide. The ligase may be split into two polypeptides bound together. The first polypeptide may include an N-terminal portion of the ligase. The first polypeptide may include a C-terminal portion of the ligase. The second polypeptide may include the N-terminal portion of the ligase. The second polypeptide may include the C-terminal portion of the ligase. The first or second polypeptide comprising a part of the ligase may be fused with at least part, or the whole, of the endonuclease.

[0090] Examples of DNA ligases are hLIG1, T4 ligase, T7 ligase, and ligases from *Aquifex aeolicus* VF5, *Neisseria meningitidis* serogroup A strain Z2491, *Neisseria meningitidis*

serogroup B strain MC58, *Pseudomonas aeruginosa* PA01, *Vibrio cholerae* El Tor N1696, *Vaccinia* virus, and *Emiliana huxleyi* virus.

[0091] The ligase may comprise a ligase that can ligate a substrate comprising DNA. In some aspects, the ligase comprises a ligase that can ligate a substrate comprising a DNA splint. For example, a DNA ligase may ligate a 5' phosphate to a 3' hydroxyl of two DNA strands that are hybridized to another DNA strand. The splinting DNA strand may include an RNA portion. For example, a DNA ligase may ligate a 5' phosphate to a 3' hydroxyl of two DNA strands that are hybridized across from a DNA portion of an RNA/DNA hybrid strand. In some aspects, the ligase comprises a ligase that can ligate a substrate comprising a DNA/RNA. In some aspects, the ligase comprises a ligase that can ligate a substrate comprising a RNA splint. For example, a DNA ligase may ligate a 5' phosphate to a 3' hydroxyl of two DNA strands that are hybridized to an RNA strand. The RNA strand may include a DNA portion. For example, a DNA ligase may ligate a 5' phosphate to a 3' hydroxyl of two DNA strands that are hybridized across from an RNA portion of an RNA/DNA hybrid strand.

[0092] In some aspects, the ligase described herein comprises any one of the ligase in **Table 7**. In some aspects, the ligase described herein comprises a polypeptide sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of the ligase in **Table 7**.

Table 7. Non-limiting examples of ligase polypeptide sequence

Name	Ligase polypeptide sequence	SEQ ID NO:
splintR (chlorella virus DNA ligase PBCV1)	MAITKPLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLS RTFKPIRNSVMNRLLELLPEGSDEISIEGATFQDTS MTGHKMYNAKFSYYWFDYVTDDPLKKYIDRVDMKNYI TVHPHILEHAQVKIIPVEINNITELLQYERDVLSKGFEG VMIRKPDGKYKFGRLKEGILLKMKQFKDAEATIISMTA LFKNTNTKTKDNFGYSKRSTHSGKVEEDVMGSIEVDYD GVVFSIGTGFDADQRRDFWQNKESYIGKMVKFKYFEMGS KDCPRFPVFIGIRHEEDR	55
hLIG1	MQRSIMSFFHPKKEGKAKKPEKEASNSSRETEPPPKAALK EWNGVVSESDSPVKRPGRKAARVLGSEGEEDDEALSPAK GQKPALDCSQVSPRPATSPENNASLSDTSPMDSSPSGIPK RRTARKQLPKRTIQEVLEEQSEDEDREAKRKKEEEEETP KESL TEAEVATEKEGEDGDQPTTPPKPLKTSKAETPTESVS EPEVATKQELQEEEEQTKPPRRAPKTLSSFFTTPRKP AVKKE VKEEPEGAPGKEGAAEGPLDPSGYNPAKNNYHPVEDACW KPGQKVPYLAVARTFEKIEEV SARLRMVETLSNLLRSVVA LSPDLLPVLYLSNLHGLPQQGLELGVGDGVLLKAVAQA TGRQLESVRAEAAEKGDVGLVAENSRSTQRLMLPPPPLTA SGVFSKFRDIARLTGSASTAKKIDIIKGLFVACRHSEARFIA RSLSGRLRLGLAEQSVLAALSQAVSLTPPGQEFPPAMVDA	56

Name	Ligase polypeptide sequence	SEQ ID NO:
	GKGKTAEARKTWLEEQGMILKQTFCEVPDLDRIPVLEH GLERLPEHCKLSPGIPLKPM LAHPTRGISEVLKRFEAAFT CEYKYDGQRAQIHALEGGEVKIFSRNQEDNTGKYPDIISRI PKIKLPSVTSFILDTEAVAWDREKKQIQPFQVLTTRKRKEV DASEIQVQVCLYAFDLIYLN GESLVREPLSRRRQLLRENFV ETEGEFVFATSLDTKDIEQIAEFLEQSVKDSCEGLMVKTLD VDATYEIAKRSHNWLK LKKDYLDGVGDTL DLVVIGAYLG RGKRAGRYGGFLLASYDEDSEELQAICKLGTGFSDEELEE HHQSLKALVLPSPRPYVRIDGAVIPDHWLDPSAVWEVKC ADLSLSPIYPAARGLVSDKGISLRFPRFIRVREDKQPEQAT TSAQVACLYRKQSQIQNQGGEDSGSDPEDTY	
hLIG1(233-919)	TPRKPAVKKEVKEEPEGAPGKEGAAEGPLDPSGYNPAKN NYHPVEDACWKPGQKVPYLAVARTFEKIEEV SARLRMVE TSLNLLRSVVALSPPDLLPVL YLSLNLHLGPPQQGLELGVG DGVLLKAVAQATGRQLESVRAEAAEKGDVGLVAENSRST QRLMLPPPPLTASGVFSKFRDIARLTGSASTAKKIDIKGLF VACRHSEARFIARSLSGRLRLGLAEQSVLAALSQAVSLTPP GQEFPPAMVDAGKGKTAEARKTWLEEQGMILKQTFCEVP DLDRIPVLEHGLERLPEHCKLSPGIPLKPM LAHPTRGISE VLKRFEAAFTCEYKYDGQRAQIHALEGGEVKIFSRNQED NTGKYPDIISRI PKIKLPSVTSFILDTEAVAWDREKKQIQPF QVLTTRKRKEVDASEIQVQVCLYAFDLIYLN GESLVREPLS RRRQLLRENFVETEGEFVFATSLDTKDIEQIAEFLEQSVKD SCEGLMVKTLDVDATYEIAKRSHNWLK LKKDYLDGVGDTL DLVVIGAYLGRGKRAGRYGGFLLASYDEDSEELQAICK LGTGFSDEELEEHHQSLKALVLPSPRPYVRIDGAVIPDHWL DPSAVWEVKCADLSLSPIYPAARGLVSDKGISLRFPRFIR VREDKQPEQATTSAQVACLYRKQSQIQNQGGEDSGSDPE DTY	57
hLIG1(119-919)	PKRRTARKQLPKRTIQEVLEEQSEDEDREAKRKKEEEEE TPKESL TEAEVATEKEGEDGDQPTTPPKPLKTSKAETPTES VSEPEVATKQELQEEEEQTKPPRRAPKTLSSFFTPRKPAVK KEVKEEPEGAPGKEGAAEGPLDPSGYNPAKN NYHPVEDA CWKPGQKVPYLAVARTFEKIEEV SARLRMVETLSNLLRSV VALSPPDLLPVL YLSLNLHLGPPQQGLELGVGDGVLLKAVA QATGRQLESVRAEAAEKGDVGLVAENSRSTQRLMLPPPPL TASGVFSKFRDIARLTGSASTAKKIDIKGLFVACRHSEARF IARSLSGRLRLGLAEQSVLAALSQAVSLTPPGQEFPPAMVD AGKGKTAEARKTWLEEQGMILKQTFCEVPDLDRIPVLE HGLERLPEHCKLSPGIPLKPM LAHPTRGISEVLKRFEAAFT TCEYKYDGQRAQIHALEGGEVKIFSRNQEDNTGKYPDIISR IPKIKLPSVTSFILDTEAVAWDREKKQIQPFQVLTTRKRKE VDASEIQVQVCLYAFDLIYLN GESLVREPLSRRRQLLRENF VETEGEFVFATSLDTKDIEQIAEFLEQSVKDSCEGLMVKTLD VDATYEIAKRSHNWLK LKKDYLDGVGDTL DLVVIGAYL GRGKRAGRYGGFLLASYDEDSEELQAICKLGTGFSDEELEE HHQSLKALVLPSPRPYVRIDGAVIPDHWLDPSAVWEVKC ADLSLSPIYPAARGLVSDKGISLRFPRFIRVREDKQPEQAT TSAQVACLYRKQSQIQNQGGEDSGSDPEDTY	58

Name	Ligase polypeptide sequence	SEQ ID NO:
hLIG3 isoform 1	MSLAFKIFFPQTLRALSRKELCLFRKHHWRDVRQFSQWSE TDLLHGHPLFLRRKPVL SFQGSHLRSRATYL VFLPGLHVG LCSGPCEMAEQRFCDYAKRGTAGCKKCKEKIVKGVCRI GKVVPNPFSESGDMKEWYHIKCMFEKLERARATTKKIE DLTELEGWEELEDNEKEQITQHIADLSSKAAGTPKKKAVV QAKLTTTGQVTSPVKGASFVTSTNPRKFSGFSKPNNSGE APSSPTPKRSLSSSKCDPRHKDCLLREFRKL CAMVADNPS YNTKTQIIQDFLRKGSAGDGFHGDVYLT VKLLLPGVIKTV YNLNDKQIVKLFSRIFNCNPDDMARDLEQGDVSETIRVFFE QSKSFPPAAKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDI ASRCTANDLKCIRLIKHDLMNSGAKHVLDALDPNAYEA FKASRNLDVVERVLHNAQEVEKEPGQRRALSVQASLMT PVQPMLAEACKSVEYAMKKCPNGMFSEIKYDGERVQVH KNGDHFYSYFSRSLKPVLPHKVAHF KDYIPQAFPGGHSMIL DSEVLLIDNKTGKPLPFGTLGVHKKAAFQDANVCLFVFDC IYFNDVSLMDRPLCERRKFLHDNMVEIPNRIMFSEMKRVT KALDLADMITRVIQEGLEGLVKDVKGT YEPGKRHWLKV KKDYLNEGAMADTADLVVLGAFYGGQSGKGGMMSIFLMG CYDPGSQKWCTVTKCAGGHDDATLARLQNELDMVKISK DPSKIPSWLKVNKIYYPDFIVDPKKA AVWEITGAEFSEKSE AHTADGISIRFPRCTIRDDKDWKSATNLPQLKELYQLSKE KADFTVVAGDEGSSTTGGSSSEENKGPSGSAVSRKAPSKPS ASTKKAEGKLSNSNSKDGNMQTAKPSAMKVGEKLATKSS PVKVGEKRKAADETLCQTKVLLDIFTGVRL YLPPSTPDFSR LRRYFVAFDGDVLVQEFDMTSATHVLGSRDKNPAAQQVSP EWIWACIRKRRLVAPC	59
hLIG3 isoform 2	MSLAFKIFFPQTLRALSRKELCLFRKHHWRDVRQFSQWSE TDLLHGHPLFLRRKPVL SFQGSHLRSRATYL VFLPGLHVG LCSGPCEMAEQRFCDYAKRGTAGCKKCKEKIVKGVCRI GKVVPNPFSESGDMKEWYHIKCMFEKLERARATTKKIE DLTELEGWEELEDNEKEQITQHIADLSSKAAGTPKKKAVV QAKLTTTGQVTSPVKGASFVTSTNPRKFSGFSKPNNSGE APSSPTPKRSLSSSKCDPRHKDCLLREFRKL CAMVADNPS YNTKTQIIQDFLRKGSAGDGFHGDVYLT VKLLLPGVIKTV YNLNDKQIVKLFSRIFNCNPDDMARDLEQGDVSETIRVFFE QSKSFPPAAKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDI ASRCTANDLKCIRLIKHDLMNSGAKHVLDALDPNAYEA FKASRNLDVVERVLHNAQEVEKEPGQRRALSVQASLMT PVQPMLAEACKSVEYAMKKCPNGMFSEIKYDGERVQVH KNGDHFYSYFSRSLKPVLPHKVAHF KDYIPQAFPGGHSMIL DSEVLLIDNKTGKPLPFGTLGVHKKAAFQDANVCLFVFDC IYFNDVSLMDRPLCERRKFLHDNMVEIPNRIMFSEMKRVT KALDLADMITRVIQEGLEGLVKDVKGT YEPGKRHWLKV KKDYLNEGAMADTADLVVLGAFYGGQSGKGGMMSIFLMG CYDPGSQKWCTVTKCAGGHDDATLARLQNELDMVKISK DPSKIPSWLKVNKIYYPDFIVDPKKA AVWEITGAEFSEKSE AHTADGISIRFPRCTIRDDKDWKSATNLPQLKELYQLSKE KADFTVVAGDEGSSTTGGSSSEENKGPSGSAVSRKAPSKPS	60

Name	Ligase polypeptide sequence	SEQ ID NO:
	ASTKKAEGKLSNSNSKDGNMQTAKPSAMKVGEKLATKSS PVKVGEKRKAADETLQCQTKRRPASEQRGRTVPAGRR	
hLIG3 isoform 3	MAEQRFCVDYAKRGTAGCKKCKEKIVKGVCRIQKVVNP FSESGGDMKEWYHIKCMFEKLERARATTKKIEDLTELEG WEELEDNEKEQITQHIADLSSKAAGTPKKKAVVQAKLTTT GQVTSPVKGASFVTSTNPRKFSGFSAKPNNSGEAPSSPTPK RSLSSSKCDPRHKDCLLREFRKLCAVADNPSYNTKTQII QDFLRKGSAGDGFHGDVYLTVKLLLPVGIKTVYNLNDKQ IVKLFSRIFNCNPDDMARDLEQGDVSETIRVFFEQSKSFPPA AKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDIASRCTAND LKCIRLIKHDLMNSGAKHVLDALDPNAYEAFKASRNQ DVVERVLHNAQEVEKEPGQRRALSVQASLMTPVQPMLE ACKSVEYAMKKCPNGMFSEIKYDGERVQVHKNGDHF SYF SRSLKPVLPHKVAHF KDYIPQAFPGGHSMILDSEVLLIDNK TGKPLPFGTLGVHKKAAFQDANVCLFVFD CIYFNDVSLM DRPLCERRKFLHDNMVEIPNRIMFSEMKRVTKALDLADMI TRVIEGLEGLVLKDVKGTYEPGKRHWLVKKDYLNEG AMADTADLVVLGAFYGGQSKGGMMSIFLMGCYDPGSQK WCTVTKCAGGHDDATLARLQNELDMVKISKDPSKIPSWL KVNKIYYPDFIVPDPKKA AVWEITGAEF SKSEAHTADGISI RFPRCTRIRDDKDWKSATNLPQLKEL YQLSKEKADFTVVA GDEGSSTTGSSEENKGPSGSAVSRKAPSKPSASTKKAEG KLSNSNSKDGNMQTAKPSAMKVGEKLATKSSPVKVGEKR KAADETLQCQTKVLLDIFTGVRLYLPPSTPDFSRLRRYFVAF DGD LVQEFDMTSATHVLGSRDKNPAAQQVSPEWIWACIR KRRLVAPC	61
hLIG3 isoform 4	MAEQRFCVDYAKRGTAGCKKCKEKIVKGVCRIQKVVNP FSESGGDMKEWYHIKCMFEKLERARATTKKIEDLTELEG WEELEDNEKEQITQHIADLSSKAAGTPKKKAVVQAKLTTT GQVTSPVKGASFVTSTNPRKFSGFSAKPNNSGEAPSSPTPK RSLSSSKCDPRHKDCLLREFRKLCAVADNPSYNTKTQII QDFLRKGSAGDGFHGDVYLTVKLLLPVGIKTVYNLNDKQ IVKLFSRIFNCNPDDMARDLEQGDVSETIRVFFEQSKSFPPA AKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDIASRCTAND LKCIRLIKHDLMNSGAKHVLDALDPNAYEAFKASRNQ DVVERVLHNAQEVEKEPGQRRALSVQASLMTPVQPMLE ACKSVEYAMKKCPNGMFSEIKYDGERVQVHKNGDHF SYF SRSLKPVLPHKVAHF KDYIPQAFPGGHSMILDSEVLLIDNK TGKPLPFGTLGVHKKAAFQDANVCLFVFD CIYFNDVSLM DRPLCERRKFLHDNMVEIPNRIMFSEMKRVTKALDLADMI TRVIEGLEGLVLKDVKGTYEPGKRHWLVKKDYLNEG AMADTADLVVLGAFYGGQSKGGMMSIFLMGCYDPGSQK WCTVTKCAGGHDDATLARLQNELDMVKISKDPSKIPSWL KVNKIYYPDFIVPDPKKA AVWEITGAEF SKSEAHTADGISI RFPRCTRIRDDKDWKSATNLPQLKEL YQLSKEKADFTVVA GDEGSSTTGSSEENKGPSGSAVSRKAPSKPSASTKKAEG KLSNSNSKDGNMQTAKPSAMKVGEKLATKSSPVKVGEKR KAADETLQCQTKRRPASEQRGRTVPAGRR	62

Name	Ligase polypeptide sequence	SEQ ID NO:
hLIG4	MAASQTSQTVASHVPFADLCSTLERIQKSKGRAEKIRHFR EFLDSWRKFHFDALHKNHKDVTDSFYFAMRLILPQLERER MAYGIKETMLAKLYIELLNLPDGGKDALKLLNYRTPGTGTH GDAGDFAMIAFYVLKPRCLQKGSLLTIQQVNDLLDSIASNN SAKRKDLIKKSLQLITQSSALEQKWLIRMIKDLKLGVSQ QTIFSVFHNDAAELHNVTTDLEKVCRLHDPVGLSDISIT LFSAFKPMLAAIADIEHIEKDMKHQSFYIETKLDGERMQM HKDGDVYKYFSRNGYNYTDQFGASPTGSLTPFIHNAFKA DIQICILDGEMMAYNPNTQTFMQGKTKFDIKRMVEDSDLQ TCYCVFDVLMVNNKKLGHETLRKRYEILSSIFTPIPGRIEIV QKTQAHTKNEVIDALNEAIDKREEGIMVKQPLSIYKPKDR GEGWLKIKPEYVSGLMDELDILVGGYWGKGSRGGMMSH FLCAVAEKPPPGEKPSVFHTLSRVGSGCTMKELYDLGLKL AKYWKPFHRKAPPSSILCGTEKPEVYIEPCNSVIVQIKAAEI VPSDMYKTGCTLRFPRIEKIRDDKEWHECMTLDDLEQLRG KASGKLASKHL YIGGDDEPQEKKRKAAPKMKKVIGIIEHL KAPNL TNVNKISNIFEDVEFCVMSGTDSQPKPDLENRIAEF GGYIVQNP GPDTYC VIAGSENIRVKNILSNKHDDVVKPAW LLECFKTKSFVPWQPRFMIHMCPTKEHFAREYDCYGDSY FIDTDLNQLKEVFSGIKNSNEQTPEEMASLIADLEYRYSWD CSPLSMFRRHTVYLD SYAVINDLSTKNEGTRLAIKALELRF HGAKVVVSCLAEGVSHVIIGEDHSRVADFKAFRRTFKRKFK ILKESWVTDSIDKCELQEENQYLI	63
T4 Ligase	MILKILNEIASIGSTKQKQAILEKNKDNELLKRVYRLTYSR GLQYYIKKWP KPGIATQSFGMLTLTDMLD FIEFTLATRKL TGNA AIEELTGYITDGKKDDVEVLRRVMMRDLECGASVSI ANKVWPGLIPEQPQMLASSYDEKGINKNIKFPFAAQLKAD GARCF AEVRGDELDDVRLLSRAGNEYLGLDLLKEELIKM TAEARQIHPEGVLIDGELVYHEQVKKEPEGLDFLFDAYPE NSKAKEFAEVAESRTASNGIANKSLKGTISEKEAQCMKFQ VWDYVPLVEIYSLPAFRLKYDVRFSKLEQMTSGYDKVILI ENQVNNLDEAKVIYKKYIDQGLEGIILKNIDGLWENARS KNLYKFKEVIDVDLKIVGIYPHRKDPTKAGGFILESECGKI KVNAGSGLKDKAGVKSHELDRTRIMENQNYIYGKILECEC NGWLKSDGRTDYVKLFLPIAIRLREDKTKANTFEDVFGDF HEVTGL	64
T7 Ligase	MMNIKTNPFAKVSFVESAIKKALDNAGYLIAEIKYDGVRG NICVDNTANSYWLSRVSKTIPALEHLNGFDVRWKRLND DRCFYKDGFM LDGELMVKGVD FNTGSGLLR TKWTDTKN QEFHEELFVEPIRKKDKVPFKLHTGHLHIKLYAILPLHIVES GEDCDVMTLLMQEHVKNMLPLLQEYFPEIEWQAAESYEV YDMVELQQLYEQKRAEGHEGLIVKDP MCIYKRGKKS GW WKMKPENEADGIIQGLVWG TKGLANEGKVIGFEVLLESG RLVNATNISRALMDEFTETVKEATLSQWGFFSPYIGDND ACTINPYDGWACQISYMEETPDGSLRHPSFVMFRGTEDNP QEKM	65

Name	Ligase polypeptide sequence	SEQ ID NO:
Taq Ligase	MTLEAARRRVNELRDLIRYHNYLYYVLDAPEISDAEYDRL LRELKELEERFPELQSPDSPTEQVGARPLESTFRPVRHPTR MYSLDNAFSLDEVRAFEERIERALGRKGPFLYTVEHKVDG LSVNLYYEEGILVFGATRGDGETGEEVTQNLITRTIPRRL TGVPDRLEVRGEVYMPIEAFRLRNQEELEAGERIFKNPRN AAAGSLRQKDPRTARRGLRATFYALGLGLEETGLKSQH DLLLWLRERGFVVEHGFTRALGAEGVEEVYQAWLKERRK LPFEADGVVVKLDDLALWRELGYTARAPRFALAYKFP AEKETRLLSVAFQVGRITGRITPVGVLEPVFIEGSEVSRVTLH NESFIEELDVRIGDWVLVHKAGGVPEVLRVLKERRTGEE KPILWPENCPECGHALIKEGKVHRCNPPLCPAKRFEAIRHY ASRKAMDIQGLGEKLEKLEKGLVRDVADLYRLKKEDL VNLERMGEKSAENLLRQIEESKGRGLERLLYALGLPGVGE VLARNLALRFGHMDRLLLEAGLEDLLEVEGVGELTARAIL NTLKDPEFRDLVRRLLKEAGVEMEAKEREAGEALKGLTFVIT GELSRPREEVKALLRRLGAKVTDSVSRKTGFLVVGENPGS KLEKARALGVPTLSEEELRYLIEERTGKDPRALTA	66
T3 Ligase	MNIFNTNPFKAVSFVESAVKKALETSGYLIADCKYDGVVRG NIVVDNVAEAAWLSRVSKFIPALEHLNGFDKRWQQLND DRCIFPDGFMLDGLMVKGVDFTGSGLLRTKWVKRDN MGFHLTNVPTKLTPKGREVIDGKFEFHLDPKRLSVRLYAV MPIHIAESGEDYDVQNLLMPYHVEAMRSLLEYFPEIEWL IAETYEYDMDSLTELYEEKRAEGHEGLIVKDPQGIYKRG KKSGWWKLKPECEADGIIQGVNWGTEGLANEGKVIGFSV LLETGRLVDANNISRALMDEFTSNVKAHGEDFYNGWACQ VNYMEATPDGSLRHPSEKFRGTEDNPQEKM	67
NAD-dependent E coli DNA ligase LigA	MESIEQQLTELRTTLRHHEYLYHVMDAPEIPDAEYDRLMR ELRELETKHPELITPDSTQRVGAAPLAAFSQRHEVPMLS LDNVFDEESFLAFNKRVDRLKNNEKVTWCCELKLDGLA VSILYENGVLVSAATRGDGTTGEDITSNVRTIRAIPLKLHG ENIPARLEVRGEVFLPQAGFEKINEDARRTGKVFANPRN AAAGSLRQLDPRITAKRPLTFFCYGVGVLEGGELPDTHLG RLLQFKKWGLPVSDRVTLCESAEEVLAFYHKVEEDRPTLG FDIDGVVIVKNSLAQQEQLGFVARAPRWAVAFKFPAQEQ MTFVRDVEFQVGRITGAIPTVARLEPVHVAGVLVSNATLH NADEIERLGLRIGDKVVIRRAGDVIPQVVNVLSERPEDTR EVVFPTHCPVCGSDVERVEGEAVARCTGGGLICGAQRKESL KHFVSRRAMDVDGMGDKIIDQLVEKEYVHTPADLFLKLT GKLTGLERMGPKSAQNVVNALEKAKETTFARFLYALGIR EVGEATAAGLAAYFGTLEALEAASIEELQKVPDVGIVVAS HVHNFFAEESNRNVISELLAEGVHWPAPIVINAEIDSPFA GKTVVLTGSLSQMSRDDAKARLVELGAKVAGSVSKKTDL VIAGEAAGSKLAKAQELGIEVIDEAEMLRLLGS	68
Thermococcus kodakarensis DNA ligase	MRYSELADLYRRLEKTTLKTTLKTKFVADFLKKTDPDELLEI VPYLILGKVFPDWERELGVGEKLLIKAVSMATGVPEKEI EDSVRDTGDLGESVALAIKKKKQKSFFSQPLTIKRVYDTF VKIAEAQGEQSQRKMKYLANLFMDAEPEEGKYLARTVL GTMRTGVAEGILRDAIAEAFRVKPELVERAYMLTSDFGYV AKIAKLEGNEGLSKVRIQIGKPIRPMLAQNAASVKDALIEM	69

Name	Ligase polypeptide sequence	SEQ ID NO:
	GGEAAFEIKYD GARVQVHKDGD KVIVYSRRL ENVT RSIPE VIEAIKAALKPEKAIVEGEL VAVGENGRPRPFQYVLRFR KYNIDEMIEKIPL ELNLFDMFVDGESLIETKFIDRRNKLEE IVKESEKIKLAEQLITKKVEEAEAFYRRAL ELGHEGLMAK RLDSIYEPGNRGKKWLKIKPTMENLDLVIIGA EWGEGRRA HLLGSFLVAAYDPHSGEFLPVGKVGS GFTDEDLVEFTKML KPYIVRQEGKFVEIEPKFVIEVTYQEIQKSPKYKSGFALRFP RYVALREDKSPEEADTI ERVAELYELQERFKAKK	
African swine fever virus DNA ligase	MLNQFPGQYSNNIFCFPIESETKSGKKASWII CVVQVQH TIIPITDEM FSTVDKDAVAEIFTKFFVEEGAVRISKMTRVTE GKNLGKKNATTVVHQAFKDALSKYNRHARQKRG AHTNR GMIPPM LVKYFNIIPKTFEEETDPIVQRKRNGVRAVACQQ GDGCILLYSRTEKEFLGLDNIKKELKQLYLFIDVRVYLDGE LYLHRKPLQWIAGQANAKTDSSELHFYVFDCFWSDQLQM PSNKRQQLLTNIFKQKEDLTFIHQVENFSVKNVDEALRLK AQFIKEGYEGAIVRNANGPYEPGYNNYHSAHLAKLKPLLD AEFILVDYTQGKKGKDLGAILWVCELPNKKRFV VTPKHLT YADRYALFQKLTPALFKKHLYGKELTVEYAELSPKTGIPL QARAVGFREPINVLEII	70
Vaccinia virus DNA ligase (strain Western Reserve)	MTSLREFRKLCCDIYHASGYKEKSKLIRDFITDRDDKYLII KLLLPGLDDRIYNMNDKQIIKLYSIIFKQS QEDMLQDLGYG YIGDTIRTFKENTEIRPRDKSILTLEDVDSFLTTLSSVTKES HQIKLLTDIASVCTCNDLKCVMMLIDKDLKIKAGPRYVLN AISPNA YDVFRKSNNLKEIENASKQNLDSISISVMTPINPM LAESCDSVNKAFFKKFPSGMFAEVKYDGERVQVHKNNNEF AFFSRNMKPVLSHKVDYLKEYIPKAFKKATSIVLDSEIVLV DEHNVPLPFGSLGIHKKKEYKNSNMCLFVFDCLYFDGFD MTDIPLYERRSFLKDVMEIPNRIVFSELTNISNESQLTDVL DDALTRKLEGLVLKDINGVYEPGKRRWLKIKRDYLNESG MADSADLVVLGAYYGKGA KGGIMAVFLMGCYDDESGK WKT VTKCSGHDDNTRLVLQDQLTMVKINKDPKKIPEWL VNKIYIPDFVVEDPKQSQIWEISGAFTSSKSHTANGISIRFP RFTRIREDKTWKESTHLNDLVNLTKS	71
Vaccinia virus DNA ligase (strain Ankara)	MTSLREFRKLCCDIYHASGYKEKSKLIRDFITDRDDKYLII KLLLPGLDDRIYNMNDKQIIKLYSIIFKQS QEDMLQDLGYG YIGDTIRTFKENTEIRPRDKSILTEEVD SFLTTLSSVTKES HQIKLLTDIASVCTCNDLKCVMMLIDKDLKIKAGPRYVLN AISP HAYDVFRKSNNLKEIENASKQNLDSISISVMTPINPM LAESCDSVNKAFFKKFPSGMFAEVKYDGERVQVHKNNNEF AFFSRNMKPVLSHKVDYLKEYIPKAFKKATSIVLDSEIVLV DEHNVPLPFGSLGIHKKKEYKNSNMCLFVFDCLYFDGFD MTDIPLYERRSFLKDVMEIPNRIVFSELTNISNESQLTDVL DDALTRKLEGLVLKDINGVYEPGKRRWLKIKRDYLNESG MADSADLVVLGAYYGKGA KGGIMAVFLMGCYDDESGK WKT VTKCSGHDDNTRLRELQDQLKMIKINKDPKKIPEWL VNKIYIPDFVVEDPKQSQIWEISGAFTSSKSHTANGISIRFP RFTRIREDKTWKESTHLNDLVNLTKS	72

Name	Ligase polypeptide sequence	SEQ ID NO:
Burkholderia pseudomallei DNA ligase	MARSPVEPPASQPAKRAAWLRAELERANYAYVLDQPDLPDAEYDRLFVELQRIEAEHPDLVTPDSPTQRVGGEAASGFTPVVHDKPMLSLNNGFADEDVIAFDKRVADGLDKATDLA GTVTEPVEYACELKFDGLAISLRYENGRFVQASTRGDGT GEDVTENIRTIRAIPLTLKGKRIPRMLDVRGEVLMFKRDFARLNERQRAAGQREFANPRNAAAGSLRQLDSKITASRPLSF FAYGIGVLDGADMPDTHSGLLDWYETLGLPVNRERAVVR GAAGLLAFFHSVGERRESLPYDIDGVVYKVNRRDEQDRL GFVSRAPRFALAHKFPAQEALTKLIAIDVQVGRTGAITPVARLEPVFVGGATVTNATLHNEDEVRRKDIRIGDTVIVRRAG DVIPEVVS AVLDRRPADAQEFVMPTECPEC GSRIERLPDEA IARCTGGLFCPAQRKQALWHFAQRRALDIDGLGEKIIDQL VEQNLVRTPADLFNLGFSTLVGLDRFAEKSARNLIDSLEK AKHTTLARFIYALGIRHVGESTAKDLAKHFGSLDPIMDAPI DALLEVNDVGPIVAESIHQFFAEHNRTVIEQLRARGKVT WPEGPPAPRAPQGVLAGKTVVLTGTLPTLTREAAKEMLE AAGAKVAGSVSKKTDYVAVAGADAGSKLAKAEELGIPVLD EAGMHTLLEGHAR	73
Alteromonas mediterranea DNA ligase	MQFFLT VFCLLLITAVTHVNAEDKLDIVDGLQLAKQYSHS RQDINIAEYWVSEKLDGIRARWDGTELRTNRNNKIAAPA WFTANWPKATIDGELWIARGQFERTASIVLSKLTSVAPHS VAGSLPRTESTVGAMTATHSLPSKRWAKIRFMAFDMPPVA GQSFD SRLNMLNNLKEATPNPTFAVVSQFTLSSVNALEEK LEQVTL SGGEGLMLHHKKAFYHSGRSDKLLKVQFEDAE AKVLAHLPGKGKFKGMMGSLLVETPAGVQFKLGTGFSEK ERQAPPAIGSWVTFKFYGVTKNGKPRFASFLRVRPPSDLP K	74
Yeast DNA ligase I (Cdc9p)	MRRLLTGCLLSSARPLKSRLPLLMSSSLPSSAGKKPKQATL ARFFTS MKNKPTEGTSPSKSSKHMLEDRMDNVS GEEY ATKKLKQTAVTHTVAAPSSMGSNFSSIPSSAPSSGVADSPQ QSQRLVGEVEDALSSNNNDHYSSNIPYSEVCFNKIEAIS SRLEIIRICSDFFIKIMKQSSKNLIPTTYLFINRLGPDYEAGLE LGLGENLLMKTISETCGKSMSQIKLKYKDIGDLGEIAMGA RNVQPTMFKPKPLTVGEVFKNLRAIAKTQGKDSQLKKMK LIKRMLTACKGIEAKFLRSLESKLRIGLAEKTVLISLSKAL LLHDENREDSPDKDVPMDVLESAQQKIRDAFCQVPNYEIV INSCLEHGIMNLDKYCTLRPGIPLKPM LAKPTK AINEVLDR FQGETFTSEYKYDGERAQVHLLNDGTMRIYSRNGENMTE RYPEINITDFIQDLDTTKNLILDCEAVAWDKDQGKILPFQV LSTRKRKDVELNDVKVKVCLFAFDILCYNDERLINKSLKE RREYLTKVTKVVPGEFYATQITTNLDELQKFLDES VNHSCEGLMVKMLEGPESHYEPSKRSRNWLKLKKDYLEGVGD SLDLCVLGAYYGRGKRTGTGGFLLGCYNQDTGEFETCC KIGTGFSDEMLQLLHDRLTPTIIDGPKATFVFDSSAEPDVW FEPTTLFEVLTADLSLSPIYKAGSATFDKGVSLRFRFLRIR EDKGVEDATSSDQIVELYENQSHMQN	75

Name	Ligase polypeptide sequence	SEQ ID NO:
Yeast DNA ligase IV	MISALDSIPEPQNFAPSPDFKWLCEELFVKIHEVQINGTAGT GKSRSFKYYEISNFVEMWRKTVGNNIYPALVLALPYRDR RIYNIKDYVLIRTICSYLKLPKNSATEQRLKDWKQRVGKG GNLSSLLVEEIAKRRAEPSSKAITIDNVNHYLDSLSGDRFA SGRGFKSLVSKSPFLHCVENMSFVELKYFFDIVLKNRVIG GQEHKLLNCWHPDAQDYLSVISDLKVVTSTKLYDPKVRLK DDDLSTKVGFAFAPQLAKKVNLSEYKICRTLHDDFLVEEK MDGERIQVHYMNYGESIKFFSRRGIDYTYLYGASLSSGTIS QHLRFTDSVKECVLDGEMVTFDAKRRVILPFGLVKGSAKE ALSFNSINNVDHFPLYMVFDLLYLNGTSLTPLPHQRKQY LNSILSPLKNIVEIVRSSRCYGVESIKKSLEVAISLGSEGVVL KYYNSSYNVASRNNNWIKVKPEYLEEFGENLDLIVIGRDS GKKDSFMLGLLVLDEEEYKKHQGDSSEIVDHSSQEKHIQN SRRRVKKILSFCSIANGISQEEFKEIDRKTRGHWKRTSEVA PPASILEFGSKIPAEWIDPSESIVLEIKSRSLDNTETNMQKY ATNCTLYGGYCKRIRYDKEWTDCTYTLNDLYESRTVKSNP SYQAERSQLGLIRKKRKRVLISDSFHQNRKQLPISNIFAGLL FYVLSDYVTEDTGIRITRAELEKTIVEHGGKLIYNVILKRHS IGDVRLISCKTTTECKALIDRGYDILHPNWVLDCAIYKRLIL IEPNYCFNVSQKMRAVAEKRVDCLGDSFENDISETKLSSL YKSQSLPMPGELEIDSEVRRFPLFLFSNRIAYVPRRKISTE DDIEMKIKLFGGKITDQQSLCNLIIPYTDPILRKDCMNEV HEKIKEQIKASDTIPKIARVVAPEWVDHSINENCQVPEEDF PVVNY	76
T6 ligase	MILKILNEIASIGSTKQKQAILEKNKDNELLKRVYRLTYSR GLQYYIKKWPKPGIATQSFGLMLTLTDMLDLFIETLATRKL TGNAAIEELTGYITDGKKDDVEVLRRVMMRDLECGASVSI ANKVWPGLIPEQPQMLASSYDEKGINKNIKFPFAAQLKAD GARCFAEVRGDELDDVRLLSRAGNEYLGLDLLKEELIKM TAEARQIHPEGVLIDGELVYHEQVKKEPEGLDFLFDHPE NSKVKDFTEVAESRTASNGIANKSLKGTISEKEAQCMKFQ VWDYVPLVEVYGLPAFRLKYDVRFSKLEQMTSGYDKVIL IENQVVNNLDEAKVIYKKYIDQGLEGIILKNIDGLWENARS KNLYKFKEVIDVDLKIVGIYPHRKDPTKAGGFILESECGKI KVNAGSGLKDKAGVKSHELDRTRIMENQNYIYGKILECEC NGWLKSDGRTDYVKLFLPIAIRLREDKTKANTFEDVFGDF HEVTGL	77
Mouse DNA ligase I	MQRSIMSFFQPTKEGKAKKPEKETPSSIREKEPPPKVALKE RNQVVPESDSPVKRTGRKVAQVLSCEGEDEDEAPGTPKV QKPVSDSEQSSPPSPDTCPENSPVFNCSPPMDISPSGFPKRR TARKQLPKRTIQDTLEEQNEDKTKTAKKRKKEEETPKESL AEAEDIKQKEEKEGDQLIVPSEPTKSPESVTLTKTENIPVCK AGVKLKPQEEEQSKPPARGAKTLSSFFTFRKPAVKTEVKQ EESGTLRKEETKGTLDPANYNPSKNNYHPIEDACWKHGQ KVPFLAVARTFEKIEEVSARLKMVETLSNLLRSVVALSPDP LLPVLYLSLNLGPPQQGLELGVGDGVLLKAVAQATGRQ LESIRAEVAEKGDVGLVAENSRSTQRLMLPPPPLTISGVFT KFCDIARLTGSASMAKKMDIIGLFFVACRHSEARYIARSLS GRLRLGLAEQSVLAALAQAVALSLTPPGQEFPTVVVDAGKG	78

Name	Ligase polypeptide sequence	SEQ ID NO:
	KTAEARKMWLEEQGMILKQTFCEVPDLDRIPVLLHEHGLE RLPEHCKLSPGVPLKPMLAHPTRGVSEVLKRFEEDFTCE YKYDGGQRAQIHVLEGGGEVKIFSRNQEDNTGKYPDIISRIPKI KHPSVTSFILDTEAVAWDREKKQIQPFQVLTTRKRKEVDA SEIQVQVCLYAFDLIYLNAGESLVRQPLSRRRQLLRENFVET EGEFVFTTSLDTKDTEQIAEFLEQSVKDSCEGLMVKTLDV DATYEIAKRSHNWLKLKKDYLDGVGDTLDLVVIGAYLGR GKRAGRYGGFLLAAYDEESEELQAICKLGTGFSDEELEEH HQSLLQALVLPTPRPYVRIDGAVAPDHWLDPSIVWEVKCA DLSLSPIYPAARGLV DKEKGISLRFPRFIRVRKDKQPEQAT TSNQVASLYRKQSQIQNQSSDLDSDVEDY	
Mouse DNA ligase IV	MASSQTSQTVAAHVPFADLCSTLERIQKGKDRAEKIRHFK EFLDSWRKFHDALHKNRKDVTDSTFYPAMRLILPQLERER MAYGIKETMLAKLYIELLNLPREGKDAQKLLNYRTPSGAR TDAGDFAMIA YFVLKPRCLQKGS LTIQQVNELLDLVASNN SGKKKDLVKKSLQLITQSSALEQKWLIRMIKDLKLGISQ QTIFSIFHNDAVELHNVTTDLEKVCRLHDPSPVGLSDISITL FSAFKPMLAAVADVERVEKDMKQQSFYIETKLDGERMQ MHKDGALYRYFSRNGYNYTDQFGESPOEGSLTPFIHNAFG TDVQACILDGEMMAYNPTTQTFMQKGKVFIDIKRMVEDS GLQTCYSVFDVLMVNKKKLGRETLRKRYEILSSTFTPIQG RIEIVQKTQAHTKKEVVDA LND AIDKREEGIMVKHPLSIY KPKDRGEGWLKIKPEYVSGLMDEL DVLIVGGYWGKGSRG GMMSHFLCAVAETPPP GDRPSVFHTLCRVGSGYTMKELY DLGLKLAKYWKPFHKKSPSSILCGTEKPEVYIEPQNSVIV QIKAAEIVPSDMYKTGSTLRFPRIEKIRDDKEWHECMTLG DLEQLRGKASGKLATKHLHVGDDEPREKRRKPISKTKK AIRIIEHLKAPNLSNVNKVSNVFEDVEFCVMSGLDGYPKA DLENRIA EFGGYIVQNP GPDTYC VIAGSENV RVKNISSDK NDVVKPEWLLCEFKTKTCVPWQPRFMIHMC PSTKQHFAR EYDCYGDSYFVDTDLDQLKEVFLGIKPSEQQTPEEMAPVI ADLECRYSWDHSPLSMFRHYTIYLDLYAVINDLSSRIEATR LGITALELRFHGAKVVSCLSEGVS HVIIGEDQRRVTD FKIF RRMLKKKFKILQESWVS DSV DKGELQEENQYLL	79
Arabidopsis DNA ligase I	MLAIRSSNYLRCIPSLCTKTQISQFSSVLISFSRQISHLRLSSC HRAMSSSRPSAFDALMSNARAAAKKKTPTTNLSRSPNK RKIGETQDANLGKTIVSEGTLPKTEDLLEPVSDSANPRSDT SSIAEDSKTGAKKAKTLSKTDEMKS KIGLLKKKPNDFDPE KMSCWEKGERVPFLFVALAFDLISNESGRIVITDILCNMLR TVIATTPEDLVATVYLSANEIAPAHEGVELGIGESTIIKAISE AFGRTEHDVKKQNTLGDGLVAKGSRSTQTM MFKPEPL TVVKVFD TFRQIAKESGKDSNEKKKNRMKALLVATTDCE PLYLTRLLQAKLRLGFSGQTVLAALGQAAVYNEEH SKPPP NTKSPLEEA AKIVKQVFTVLPVYDIIVPALLSGGVWNLPKT CNFTLGVP IGPMLAKPTKGVAEILNKFQDIVTCEYKYDG ERAQIHFMEDGTFEIYSRNAERNTGKYPDVALALSRLKKP SVKSFILDCEVVAFDREKKKILPFQILSTRARKNVNVNDIK VGVCIFAFDMLYLNGQQLIQENLKIRREKLYESFEEDPGYF QFATAVTSNDIDEIQKFLDASVDVGCEGLIKTLDS DATYE	80

Name	Ligase polypeptide sequence	SEQ ID NO:
	PAKRSNNWLKLLKDYMDSIGDSVDLVPIAAFHGRGKRTG VYGAFLLACYDVKKEEFQSICKIGTGFSAMLDERSSSLRS QVIATPKQYYRVGDSLNPVWFEPTEVWEVKAADLTISPV HRAATGIVDPDKGISLRFPRLLRVREDKKPEEATSSEQIAD LYQAQKHNPSPNEVKGDDD	
Arabidopsis DNA ligase IV	MTEEIFSVLVSLFNWIKSKTSSQKRSKFRKFLDTYCKPS DYFVAVRLIIPSLDRERGSYGLKESVLATCLIDALGISRDAP DAVRLLNWRKGGTAKAGANAGNFSLIAAEVLQRRQGMA SGGLTIKELNDLLDRLASSENRAEKTLLVSTLIQKTNAQEM KWVIRIILKDLKLGMEKSIFQEFHPDAEDLFNVTCDLKL CEKLRDRHQRHQRQDIEVGKAVRPQLAMRIGDVNAAWK KLHGKDVVAECKFDGDRIQIHKNGTDIHYFSRNFLDHSEY AHAMSDLIVQNILVDKCILDGEMLVWDTSLNRF AEFGSNQ EIAKAAREGLDSHKQLCYVAFDVLVVGDTSVIHQSLKERH ELLKKVVKPLKGRLEVLVPEGGLNVHRPSGEPSSIVVHA AADVERFFKETVENRDEGIVLKDLESKWEPGDRSGKWMK LKPEYIRAGADLDVLIIGGYYGSGRRGGEVAQFLVALADR AEANVYPRRFMSFCRVGTGLSDDELNTVVSKLKPYFRKN EHPKKAPPSFYQVTNHSKERPDVWIDSPEKSILSITSDIRTI RSEVFVAPYSLRFPRIDKVRDYDKPWHECLDVQAFVELVNS SNGTTQKQKESESTQDNPKVNKSSKRGEKKNVSLVPSQFI QTDVSDIKGKTSIFS NMIFYFVNVP RSHSLET FHKMVVENG GKFSMNLNNSVTHCIAAESSGIKYQAAKRQRDVIHFSWVL DCCSRNKMLPLLPKYFLHLDASRTKLQDDIDEFSDSYW DLGLEGLKQVLSNAKQSEDSKSIDYYKKKLCPEKRWSCLL SCCVYFYPYSQTLSTEEALLGIMAKRLMLEVLMAGGKV SNNLAHASHLVVLAMAEPLDFTLVSKSFSEMEKRLLLKK RLHVVSSSHWEESLQREEKLCEDEVYTLRPKYMESDTEES DKSEHDTTEVASQGSQAQTEPASSKIAITSSRGRSNTRAVK RGRSSTNSLQRVQRRRGKQPSKISGDETEESDASEEKVSTR LSDIAEETDSFGAQRNSSRGKCAKRGKSRVGQTQRVQRS RRGKKAAGIGDESDENDDELGNNNVSADAEENAAAGRS VENEETREPDIKAYTESQQRDNTVAVEEALQDSRNAKTE MDMKEKLQIHEDPLQAMLMKMFPIPSQKTTETSNRTTGE YRKANVSGECESSEKRKLD AETDNTSVNAGAESDVVPL VKKKKVSYRDVAGELLKDW	81
Arabidopsis DNA ligase 6	MASDSAGATISGNFSNSDNSETLNLNTTKLYSSAISSISPQF PSPKPTSSCPSIPNSKRIPNTNFIVDLFRLPHQSSSVAFFLSHF HSDHYSGLSSSWSKGIYC SHKTARLVAEILQVPSQVFAL PMNQMVKIDGSEVVLIEANHCPGAVQFLFKVKLESSGF EK YVHTGDFRFCDEMRFDPFLNGFVGCDGVFLDTTYCNPKF VFPSQEE SVGYVVSVIDKISEEKVLFLVATYVVGKEKILVE IARRCKRKIVVDARKMSMLSVLGCGEEGMFTEDENESDV HVVGWNVLGETWPYFRPNFVKMNEIMVEKGYDKVVGVFV PTGWTYEVKRNKFAVRFKDSMEIHLVPYSEHSNYDELREF IKFLKPKRVIPTVGV DIEKFDCKEVNKMQKHFSGLVDEMA NKKDFLLGFYRQSYQKNEKSDVDVVS HSAEVYEEEEKNA CEDGGENVPSSRGPI LHDTTPSSDSRLLIKLRDSLPAWVTE EQMLDLIKKHAGNPVDIVSNFYEYEAELYKQASLPTPSLN	82

Name	Ligase polypeptide sequence	SEQ ID NO:
	NQAVLFDDDDVTDLQPNPVKGICPDVQAIQKGFDLPRKMN LTKGTISPGKRGKSSGSKSNKKAKKDPKSKPVGPGQPTLF KFFNKVLDGGSNSVSVGSETEECNTDKKMVIDASEAYK EVTDOFIDIVNGSESLRDYAASIIDEAKGDISRALNIYYSKP REIPGDHAGERGLSSKTIQYPKCSEACSSQEDKKASENSGH AVNICVQTSAEESVDKNYVSLPPEKYQPKEHACWREGQP APYIHLVRTFASVESEKGGIKAMSMCNMFRSLFALSPED VLPVYLCNTKIAADHENIELNIGGSLISSALEEACGISRST VRDMYNSLGDLDGVAQLCRQTQKLLVPPPLLVRDVFST LRKISVQTGTGSTRLLKKNLIVKLMRSCREKEIKFLVRTLAR NLRIGAMLRTVLPALGRAIVMNSFWNDHNKELSESCFREK LEGVSAAVVEAYNILPSLDVVVPSLMDKDIEFSTSTLSMVP GPIKPMLAKIAKGVQEFFNLSQEKAFCEYKYDQGQRAQIH KLLDGTVCIFSRNGDETTSRFPDLVDVIKQFSCPAAEFML DAEVVATDRINGNKLMSFQELSTRERGSKDALITTESIKVE VCVFVFDIMFVNGEQLLALPLRERRRRLKEVFPETRPGYL EYAKEITVGAEELNNHDTLSRINAFLEEFQSSCEGIMV KSLDVNAGYCPTKRSDSWLKVKRDYVDGLGDTLDLVP AWYGNRKGAGWYSPFLMACFNPETEEFQSVCRVMSGFS DAFYIEMKEFYSEDKILAKKPPYYRTGETPDMWFSAEVV WEIRGADFTVSPVHSAASLGLVHPSRGISVRFPFISKVTDR NPEECSTATDIAEMFHAQTRKMNITSQH	
Bacillus subtilis DNA ligase	MDKETAKQRAEELRTINKYSYEEYTLDEPSVPDAEYDRL MQELIAIEEHPDLRTPDSPTQRVGGAVLEAFQKVTHGTP MLSLGNFADDLRDFDRRVRQSVGDDVAYNVELKIDGL AVSLRYEDGYFVRGATRGTGTTGEDITENLKTIRNIPKLM NRELSIEVRGEAYMPKRSFEALNEERIKNEEPPFANPRNAA AGSLRQLDPKIAAKRNLDIFVYSIAELDEMGMVETQSQGLD FLDELGFKTNQERKKCGSIEEVITLIDELQAKRADLPYEID GIVIKVDSLQDQEEELGFTAKSPRWAIAYKFPAAEEVVTCLL DIELNVGRTGVITPTAILEPVKVAGTTVSRASLHNEDLIKE KDIRILDKVVVKKAGDIPEVVNVLDVQRTGEEKEFSMPTE CPECGSELVRIEGEVALRCINPECPAQIREGLIHVSRNAM NIDGLGERVITQLFEENLVRNVADLYKLTKEVVIQLERMG EKSTENLISSIQSKENSLERLLFGLGIRFIGSKAAKTLAMH FESLENLKKASKEELLAVDEIGEKMADAVITYFHKEEMLE LLNELQELGVNTLYKGPKKVKAEDSDSYFAGKTIVLTGKL EELSRNEAKAQIEALGGKLTGSVSKNTDLVIAGEAAGSKL TKAQELNIEVWNEEQLMGELKK	83
Bacillus stearothermophilus	MDRQQAERRAAELRELLNRYGYEYVLDPRSPDAEYDR LMQELIAIEEQYPELKTSDSPTQRIGGPPLEAFRKVAHRVP MMSLANAFGEGLRDFDRRVRQEVGEAAVYVCELKIDGLA VSVRYEDGYFVQGATRGTGTTGEDITENLKTIRSLPLRLK EPVSLARGEAFMPKASFLRLNEERKARGEELFANPRNAA AGSLRQLDPKVAASRQLDLFVYGLADAEALGIASHSEALD YLQALGFKVNPERRRCANIDEVIAFVSEWHDKRPQLPYEI DGIVIKVDSFAQQRALGATAKSPRWAIAYKFPAAEEVVTTLI GIEVNVGRTGVVTPTAILEPVRVAGTTVQRTLHNEDFIRE KDIRIGDAVIIKKAGDIPEVVGVVVDRRDGDETPFAMPTH	84

Name	Ligase polypeptide sequence	SEQ ID NO:
	CPECESELVRLEGEVALRCLNPNCPAQLRERLIHFASRAA MNIEGLGEKVVTQLFNAGLVRDVADLYCLTKEQLVGLER MGEKSAANLLAAIEASKQNSLERLLFGLGIRYVGAKAAQL LAEHFETMERLERATKEELMAVPEIGEKMADAITAFFAQP EATELLQELRAYGVNMAYKGPKRSAEAPADSAFAGKTVV LTGKLASMSRNEAKEQIERLGGRVTGSVSRSTDVIAGED AGSKLEKAQQLGIEIWDESRLFQEI NRGKR	
Haemophilus influenzae Rd	MKFYRTLLFFASSFAFANS DLM LHTYNNQPIEGWVMSE KLDGVRGYWNGKQLLTRQGQRLSPPAYFIKDFPPFAIDGE LFSE RNFEEISTITKSFKGDGWEK LKLYVFDVPDAEGNLF ERLAKLKAHLLEHPTTYIEIIEQIPVKDKTHLYQFLAQVEN LQGE GVVVRNP NPYERKRSSQILKLKTARGEECTVIAHH KGKGQFENVMGAL TCKNHRGEFKIGSGFNLNERENPPPIG SVITYKYRGITNSGKPRFATYWREKK	85
Pseudoalteromonas haloplanktis	MSSSISEQVNHRLRIILEQHNYNYVLDTPSIDSEYDRLLRE LSALETEHPEFLTADSPTQKVGG AALSKFEQVAHQVPMLS LDNAFSEEEFTAFNRRIKERLMSTDELTCCEPKLDGLAVS IYRDGVLVQAATR GDGFTGENITQNVKTIRNVPLKLRGD YPKELEVRGEVFMDSAGFDKLNTEAEKRGEKVFNPRNA AAGSLRQLDSKITAKRPLMFYAYSTGLVADGNIPEDHYQQ LEKLTDWGLPLCPETKLVEGPKAALAYYRDILTRSELKY EIDGVVIKINQKTLQERLGFVARAPRWAIAYKFPAQEEITQ LLDVDFQVGRGTGAITPVARLKPVFVGGVTVSNATLHNSDE VARLGVKVGD TVIIRRAGDVIPQITQVVLERRPDDARDIEF PTTCPICDSHVEKVEGEAVARCTGGLVCPAQRKQAIKHFA SRKALDIDGLGDKIVDQLVDRELKTPADLFILKQGHFESL ERMGPKS AKNLVTALEEAKGTTLAKFLYSLGIREAGEATA QNLANHFLT ENVINASIDSLTQVSDVGEIVA AHVRGFFDE EHNLA VVNALIDQGVNWPALSAPSEEEQPLAGLTYVLTGT LNTLNRNDAKARLQQLGAKVSGSVSAKTDALVAGEKAG SKLTKAQDLGIDILTEDELIELLIKHN G	86
Rhodothermus marinus	METHTAPQTAEARLLEATH TLLQTVRQRDLEAIDRKEAEA LAARLREVLNQHAYRYVLDNPLIPDADYDLLMQALRKL EARFPELVTPDSPTQRVGGPPLGRFEKVRHPELLSLNNAF GEEDVRVWYERCCRMLAERLGQPVPQPAVTAELKIDGLA MALTYENGVL SVGATRGDGIEGENVTQNVRTIPAIPLRIPV DPAVGPPPTRLEVRGEVYMRKRDFERLNEQLQARGERPF ANPRNAAAGSVRQLNPQVTALRPLSFFAYGIGPVEGAEP DSQYEV LQWLGR LGFPVNEHARRFEHLDDVLEYCRYWTE HRDEL DYEIDGVVLKIDHRPWQALLGAISNAPRWAVAYK FPAREAITRLLDIMVSVGR TG VVKPVAVLEPVEVGGVTVS QATLHNEDYVRSRDIRIGDLVVIRAGDVIPQVVRPVVEA RTGNERPWRMPERCPSCGSQLVRLPGEADYYCVASDCPA QFVRLLEHFAGRDAMDIEGMGSQVARQLAESGLVRPLSD LYRLKLEDLLKLEGF AETRARNLLRAIEASKQRPLSRLLFG LGIRHV GKT TAELLVQRFASIDELAAATIDELAALEGVGPI TAESIANWFRVEDNRRLLIEELKELGVNTQRLPEEAPAAESP	87

Name	Ligase polypeptide sequence	SEQ ID NO:
	VRGKTFVLTGALPHLTRKEAEELIKRAGGRVASSVSRNTD YVVVGENPGSKYDRARQLGIPMLDEDGLLRLLGMK	
Thermus filiformis	MTREEARRRINELRDLIRYHNYRYVVLADPEISDAEYDRL LRELKELEERFPEFKSPDSPTEQVGARPLEPTFRPVRHPTR MYSLDNAFTYEEVLAFEERLERALGRKRPFLYTVEHKVD GLSVNLYYEEGVLVFGATRGDGEVGEVVTQNLLTIPTIPR LKGVPDRLEVRGEVYMPIEAFRLRLNEELEERGEKVFKNPR NAAAGSLRQKDPRVTAKRGLRATFYALGLGLEESGLKSQ YELLWLKEKGFPVEHGYEKALGAEGVEEVYRRFLAQRH ALPFEADGVVVKLDDLALWRELGYTARAPRFALAYKFPA EEKETRLLDVVFQVGRTGRVTPVGVLEPVFIEGSEVSRVTL HNESYIEELDIRIGDWVLVHKAGGVIPEVLRVLKERRTGEE RPIRWPETCPECGHRLVKEGKVHRCNPPLCPAKRFEAIRH YASRKAMDIEGLGEKLIERLLEKGLVRDVADLYHLRKEDL LGLERMGEKSAQNLLRQIEESKHRLGLERLLYALGLPGVGE VLARNLARRFGTMDRLLEASLEELLEVEEVGELTARAILE TLKDPAFRDLVRRLKEAGVSMESKEEVSDLLSGLTFVLTG ELSRPREEVKALLQRLGAKVTDSVSRKTSYLVVGENPGSK LEKARALGVAVLTEEEFWRFLKEKGAPVPA	88
Thermus scotoductus	MTLEEARKRVNELRDLIRYHNYRYVVLADPEISDAEYDRL LRELKELEERFPELKSPPSPTEQVGAKPLEATFRPIRHPTR MYSLDNAFNFDELKAFEERIGRAGREGPFAYTVEHKVD GLSVNLYYEDGVLVWGATRGDGEVGEVVTQNLLTIPTIPR RVKGVPERLEVRGEVYMPIEAFRLRLNEELEEKGEKIFKNPR NAAAGSLRQKDPRITARRGLRATFYALGLGLEESGLKTQL DLLHWLREKGFPVEHGFARAEGAEGVERIYQGWLKERRS LPFEADGVVVKLDELALWRELGYTARAPRFAYKFPAAEE KETRLQVVFQVGRTGRVTPVGILEPVFIEGSVVSRLVTLHN ESYIEELDVRIGDWVLVHKAGGVIPEVLRVLKEKRTGEER PIRWPETCPECGHRLVKEGKVHRCNPPLCPAKRFEAIRHY ASRKAMDIGGLGEKLIEKLLEKGLVKDVADLYRLKKEDL LGLERMGEKSAQNLLRQIEESKGRGLERLLYALGLPGVGE VLARNLAHFMTMDRLLEASLEELLQVEEVGELTARGIYE TLQDPAFRDLVRRLKEAGVVMEAKERGEEALKGLTFVIT GELSRPREEVKALLRRLGAKVTDSVSRKTSYLVVGENPGS KLEKARALGVPTLTEEELYRLIEERTGKPVETLAS	89

Name	Ligase polypeptide sequence	SEQ ID NO:
Thermus species AK16D	MTLEEARRRVNELRDLIRYHNYLYYVLDAPESDAEYDRL LRELKELEERFPELKSPDSPTEQVGARPLEATFRPVRHPTR MYSLDNAFSLDEVRAFEERIERALGRKGPFLYTVERKVDG LSVNLYYEEGILVFGATRGDGETGEEVTQNLLTIPTIPRRLT GVPDRLEVRGEVYMPIEAFRLRNQELEEAGERIFKNPRNA AAGSLRQKDPRVTARRGLRATFYALGLGLEETGLKSQHD LLWLRERGFVVEHGFTRALGAEGVEEVYQAWLKERRKL PFEADGVVVKLDDLALWRELGYTARTPRFALAYKFPAEE KETRLLSVAFQVGRTGRITPVGVLEPVFIEGSEVSRVTLHN ESFIEELDVRIGDWVLVHKAGGVIPEVLRVLKERRTGEEKP IWPENCPECGHALIEKGKVHRCNPCLCPAKRFEAIRHYAS RKAMDIQGLGEKLEKLEKGLVRDVADLYRLKKEDLVN LERMGEKSAENLLRQIEESKGRGLERLLYALGLPGVGEVL ARNLALRFGHMDRLLEAGLEDLLEVEGVGELTARAILNTL KDPEFRDLVRRLKEAGVEMEAKEREAGEALKGLTFVITGEL SRPREEVKALLRRLGAKVTDSVSRKTSFLVVGENPGSKLE KARALGVPTLSEELYRLIEERTGKDPRALTA	90
Thermus thermophilus HB8	MTLEEARKRVNELRDLIRYHNYRYVLDAPESDAEYDRL LRELKELEERFPELKSPDSPTLQVGARPLEATFRPVRHPTR MYSLDNAFNLDELKAFEERIERALGRKGPFAYTVEHKVD GLSVNLYYEEGVLVYGATRGDGEVGEVVTQNLLTIPTIPR RLKGVPERLEVRGEVYMPIEAFRLNNEELEERGERIFKNPR NAAAGSLRQKDPRITAKRGLRATFYALGLGLEEVEREGV ATQFALLHWLKEKGFPVEHGYARAVGAEGVEAVYQDWL KKRRALPFEADGVVVKLDELALWRELGYTARAPRFAIAY KFPAAEEKETRLLDVVVFQVGRTGRVTPVGILEPVFLEGSEVS RVTLHNESYIEELDIRIGDWVLVHKAGGVIPEVLRVLKERR TGEERPIRWPETCPECGRHLLKEGKVHRCNPCLCPAKRFE AIRHFASRKAMDIQGLGEKLIERLLEKGLVKDVADLYRLR KEDLVGLERMGEKSAQNLLRQIEESKKRGLERLLYALGLP GVGEVLARNLAARFGNMDRLLEASLEELLEVEEVGELTA RAILETLKDPAFRDLVRRLKEAGVEMEAKKEKGGEALKGL TFVITGELSRPREEVKALLRRLGAKVTDSVSRKTSYLVVG ENPGSKLEKARALGVPTLTEEELYRLLEARTGKKAEEELV	91
Zymomonas mobilis	MNADIDLFSYLNPEKQDLSALAPKDLREQAVIELERLAK LISHYDHLHYHDKDNPAVPDSEYDALVLRNRRIEQFFPDILIR PDSPSKKVGSRPNSRLPKIAHRAAMLSDNGFLDQDVEDF LGRVRRFFNLKENQAVICTVEPKIDGLSCSLRYEKGILTQA VTRGDGVIGEDVTPNVRVIDDIPKTLKGDWNPWEIIEIRGEV YMAKSDFAALNARQTEENKKLFANPRNAAAGSLRQLDPN ITARRSLRFLAHGWGEATSLPADTQYGMMKVIESYGLSVS NLLARADDIGQMLDFYQKIEAERADLDFDIDGVVYKLDQ LDWQQRFQFSARAPRFALAHKFPAEKAQTLLDIEIQVGR TGVLTVPVAKLEPVTVGGVVSSATLHNSDEIERLGVRPGD RVLVQRAGDVIPQIVENLTPDVDRPIWRFPHRCPCDSVA RREEGEVAWRCTGGLICPAQRVERLCHFVSRTAFEIEGLG KSHIESFFADKLIETPADIFRLFQKRQLLIEREGWGELSVDN LISAIDKRRKVPFDRFLFALGIRHVGAVTARDLAKSYQTW DNFKAAIDEAAHLRTILQPSSESEEKQYQKRVKELISFFHI	92

Name	Ligase polypeptide sequence	SEQ ID NO:
	PNMGGKIIRSLLDFFAETHNSDVVSDLLQEVQIEPLYFELASSPLSGKIIVFTGSLQKITRDEAKRQAENLGAKVASSVSKKTNLVVAGEAAGSKLSKAKELDISIIDEDRWHRIVENDGQDSIKI	
Campylobacter jejuni	MKKEEYLEKVALANLWMRAYYEKDEPLASDEEYDVLIRELRVFEEQNKDEISKDSPTQKIAPTIQSEFKKIAHLKRMWSMEDVFDESELRAWAKRAKCEKNFFIEPKFDGASLNLLYENGKLVS GATRGDGEVGEDITLNVFEIENIPKNIA YKERIEIRGEVVILKDDFEKINEKRALLNQSLFANPRNAASGSLRQLDTSITKERNLKFYPWGVGENTLNF TKHSEVMQFIRELGFLKDDFIKLCANLDEV LKAYDELLALREKKPMMMDGMVVRIDDLALCEELGYTVKFPKFMAAFKFPALEKTTRLIGVNLQVGRSGVITPVA VLEPVNLDGVVVK SATLHNFDEIARLDVKINDFVSVIRSGDVIPKITK VFKDRREGLEMEISRPKLCPTCQSELLDEGT LKICQNIDCEDRLVNSIIHFVSKKCLNIDGLGENIVELLYKHKKITTLESIFHLKFSDFEGLGFKKINNLLNAIEQARECELF RFITALGIEHIGEVAAKKLSLSFGKEWHKQSFEAYANLEGFGEQMALSLCEFTRVNHVRIDEFYKLLNLKIEKLEIKSDGVIFGKTFVITGTL SRPRDEFKALIEKLGGKVSSSVSKKTDYVLFGE EAGSKLIKAKELEVKCIDESAFNELVKE	93
Mycobacterium tuberculosis ligA	MSSPDADQTAPEVLRQWQALAEVREHQFRYYVRDAPISDAEFDELLRRLEALEEQHP ELRTPDSPTQLVGGAGFATDFEPVDHLERMLS LDNAFTADELA AWAGRIHAEVGDAAHYLCELKIDGVALSLVYREGRLTRASTRGDGR TGEDVT LNARTIADVPERLTPGDDYPVPEVLEVRGEVFFRLDDFQALNASLVEEGKAPFANPRNSAAGSLRQKDP AVTARRRLRMICHGLGHVEGFRPATLHQAYLALRAWGLPVSEHTTLATDLAGVRERIDYWGEHRHEVDHEIDGVVVKVDEVALQRRLGSTSRA PRWAIA YKYPPEEAQTKLLDIRVNVGRTGRITPFAFMTPVKVAGSTVGQATLHNASEIKRKGV LIGDTV VIRKAGDVIPEVLGPVVELRDG SEREFIMPTTCPECGSPLAPEKEGDADIRCPNARGCPGQLRERVFHVASRNGLDIEVLGYEAGVALLQAKVIADEGELFALTERDLLRTDLFR TKAGELSANGKRLLVNLDKAKAAPLWRVLVALSIRHVGPTAARALATEFGSLDAIAAAS TDQLAAVEGVGPTIAAAVTEWFAVDWHREIVDKWRAAGVRMVDERDESVPRTL AGLTIVVTGSLTGFSRDDAKEAIVARGGKAAGSVSKKTNYV VAGDSPGSKYDKA VELGVPILDEDGFRRLADGPASRT	94

Name	Ligase polypeptide sequence	SEQ ID NO:
Emiliana huxleyi virus DNA ligase	MEAMCTECEDRDARLDVIDIQLFHALNPKSCNRTTWEQV PKIMGKQGDFVAEGKLDGERDASHLYGESMEDVLCECVR EDVTSLLLDGEMMVVDLETGRYLPFGENRSLKDFGTSMR HCFVAFDLLLLYNGRSMTGATLAERSELLRKAVRTKQHAL TLIERFEVGERGAGATTAVMRQLDVMMSRGLEGVVFVKSL SSKYDPGSRDKSWIKLKPDFVDGMDTLDLLILGGYYGE GRRRSGAVSTFLMGVRAPEAAKRVGGAAHPLFYFCKV GTGYSLPQLRELRLMPASLTRRRGNALGHGASLTAVSC EQVSHEWKNSRRPAHLCHWEPSKRDDIPDYWFEPEASVV LELTAFEIITPRESFLPANYTLRFPRVKRVRYDKGWEGAET FERVVELFKECDGRLSANKRRAEIEAASRASAGPAACKRA AGVAPTVGVPWHLKLSADLANTAVECYALDGVVAVVKG TLRRPGVETQIKRLGGKVHKNMTSLTTHLVDAPGAEVL AEVERARRGGGSFEVVTAAWVDECSR VHARVTLEPRYVR HVSEATREQIEAIMDEWGDNYTIAADPESLVDSMRLVREQ RSAGGNCGDSPLAREAHVADALRDLDDETAVALTRYA MLRGVVAYVPRGSVALRLRLRLLLGAQTVDEPSADSTHAV LSASTSADERQRLRDKFTEDRVRDGRPSCGRHIVSDRWLA ECERRGQREPEAQEDAWFGDRVGIRDRL	95
Lymantria dispar multicapsid nuclear polyhedrosis virus	MENHDSFYKFCQLCQSLYDADDHQEKRDALERHFADFRG SAFMWRELLAPAESDAAADRELTLIFETILSIERTEQENV RNLKCTIDGAAVPLSRESRITVPQVYEFINDLRGSGSRQER LRLIGQFAAGCTDEDLLTVFRVVS DHAHAGLSAEDVMEL VEPWERFQKPVPPALAQPCRRLASVLVKHPEGALAEVKY DGERVQVHKAGSRFKFFSRTLKPVPEHKVAGCREHLTRAF PRARNFILDAEIVMVDGSGEALPFGTLGRLKQMEHADGH VCMYIFDCLRYNGVSYLNATPLDFRRRVLQDEIVPIEGRV VLSAMERTNTLSELRRFVHRTLATGAEGVVLKGRLESSYAP NKRDWFKMKKEHLCDGALVDTLDLVVLGAYYGTGRNC RKMSVFLMGCLDRESNVWTTVTKVHSGGLADAALTALSK ELRPLMAAPRDDLPEWFDCNESMVPHLLAADPEKMPVW EACSEMKANIGAHTAGVTMRFPVRKFRPKDKDWSTATD LQEAELIRNSQENTKKT FARLATTYDGSPNKKLKLN	96

[0093] Some aspects include a DNA ligase that ligates DNA strands base paired to a DNA splint. In some embodiments, the DNA ligase ligates DNA strands base paired to an RNA splint. In some embodiments, the DNA ligase comprises an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 55-96, or a functional fragment thereof.

[0094] In some aspects, the ligases comprises at least one NLS (e.g., any one of the NLS in **Table 2**). In some aspects, the ligase comprises at least one additional domain. In some aspects, the at least one additional domain is a dimerization domain (e.g., any one of the dimerization domain in **Table 3**). In some aspects, the ligase comprising a dimerization domain can be dimerized with an endonuclease to form a heterodimer. In some aspects, the at least one

additional domain is a functional domain. For example, the functional domain can comprises a chromatin modifying domain (e.g., any one of the chromatin modifying domain in **Table 4**) or a cell penetrating peptide (e.g., any one of the cell penetrating peptide in **Table 5**). In some aspects, the ligase comprises a linker, where the linker can covalently connect the ligase with another polypeptide (e.g., the endonuclease). In some aspects, the linker covalently connects the ligase to the at least one additional domain. In some aspects, the ligase comprises a tag (e.g., any one of the tag in **Table 6**), where the tag can be used for increasing expression, identifying, or purifying the ligase. A linker may separate the ligase from a nuclear localization signal, a chromatin modifying domain, a cell penetrating peptide, or a tag polypeptide. Any linker described herein may be included.

[0095] The ligase may comprise a binding motif for binding to a nucleic acid motif (e.g., a hairpin motif). In some aspects, the ligase (e.g. DNA ligase) comprises an MS2 coat protein (MCP) peptide. The ligase may include a hairpin binding motif such as an MCP peptide. The MCP peptide may be useful for recruiting the ligase to a guide nucleic acid comprising an MS2 hairpin. A benefit of using a MCP peptide and MS2 hairpin is to separate the ligase and endonuclease such as a Cas nickase (or a portion of them), and allow fitting within separate vectors such as AAV vectors. In some aspects, the ligase comprises a loop region. In some aspects, the loop region is a 2a loop or a 3a loop. The loop region may comprise a 2a loop. The loop region may comprise a 3a loop.

Fusion Proteins

[0096] Disclosed herein are fusion proteins. Some aspects include a nucleic acid (e.g. an expression vector) encoding a fusion protein. The fusion protein may include an endonuclease. The fusion protein may include a ligase. The fusion protein may include a linker. The endonuclease and ligase may be connected through a linker. The fusion protein may be an example of a covalently coupled endonuclease and DNA ligase. The fusion protein may comprise an endonuclease such as an RNA-guided endonuclease fused to a DNA ligase.

[0097] The fusion protein may be non-naturally occurring. The fusion protein may be engineered. The fusion protein may be synthetic. The fusion protein may be pre-synthesized. The fusion protein may be added to a subject or a cell. The fusion protein may be encoded by a nucleic acid. The encoding nucleic acid may be engineered, synthetic, or added to a subject or a cell.

[0098] The fusion protein may include one of various orientations. For example, the fusion protein may include an RNA-guided endonuclease upstream (e.g. N-terminal or in the N-direction) or downstream (e.g. C-terminal or in the C-direction) relative to the DNA ligase. The fusion protein may include an RNA-guided endonuclease amino (N)-terminal to the DNA

ligase. The fusion protein may include an RNA-guided endonuclease carboxy (C)-terminal to the DNA ligase. The endonuclease may be in the amino direction within the fusion polypeptide relative to the ligase. The endonuclease may be in the carboxy direction within the fusion polypeptide relative to the ligase. The endonuclease may be N-terminal. The endonuclease may be C-terminal. The ligase may be N-terminal. The ligase may be C-terminal.

[0099] The fusion protein may include a nuclear localization signal, chromatin modifying domain, cell penetrating peptide, tag polypeptide, or exonuclease. The fusion protein may include a nuclear localization signal. The fusion protein may include a chromatin modifying domain. The fusion protein may include a cell penetrating peptide. The fusion protein may include a tag polypeptide. The fusion protein may include an exonuclease. Any of the nuclear localization signal, chromatin modifying domain, cell penetrating peptide, tag polypeptide, or exonuclease, endonuclease, or ligase may be directly connected to another or to the endonuclease or ligase. Any of the nuclear localization signal, chromatin modifying domain, cell penetrating peptide, tag polypeptide, or exonuclease, endonuclease, or ligase may be connected by a linker to another or to the endonuclease or ligase. Multiple linkers may be included in the fusion protein. The fusion protein may exclude a polymerase.

[00100] A linker may include an amino acid linker. The amino acid linker may include a length of residues. The length may include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 residues, or a range of residues defined by any two of the aforementioned integers. The length may include at least 1 residue, at least 2 residues, at least 3 residues, at least 4 residues, at least 5 residues, at least 6 residues, at least 7 residues, at least 8 residues, at least 9 residues, at least 10 residues, at least 15 residues, at least 20 residues, at least 25 residues, at least 30 residues, at least 40 residues, at least 50 residues, at least 60 residues, at least 70 residues, at least 80 residues, at least 90 residues, or at least 100 residues. In some aspects, the length may include less than 2 residues, less than 3 residues, less than 4 residues, less than 5 residues, less than 6 residues, less than 7 residues, less than 8 residues, less than 9 residues, less than 10 residues, less than 15 residues, less than 20 residues, less than 25 residues, less than 30 residues, less than 40 residues, less than 50 residues, less than 60 residues, less than 70 residues, less than 80 residues, less than 90 residues, or less than 100 residues. Examples of residues may include alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, or valine, or any combination thereof. The linker may be non-enzymatic, or may lack any enzymatic activity.

[00101] A connection may be covalent. A covalent connection may include a peptide bond. The peptide bond may include amide bond. A connection may be between an N-terminus and

another N-terminus. A connection may be between a C-terminus and another C-terminus. A connection may be between an N-terminus and a C-terminus. A connection may be between a C-terminus and an N-terminus.

[00102] The fusion protein may include connections in various orientations. The endonuclease may be connected at its C-terminus. The endonuclease may be connected at its N-terminus. The ligase may be connected at its C-terminus. The ligase may be connected at its N-terminus.

[00103] Fig. 7 illustrates some examples of fusion protein. The figure includes examples of arrangements and orientations of the endonuclease, linker, ligase, or nuclear localization signal. Other aspects may be incorporated into the examples shown.

Non-Covalently Coupled Proteins

[00104] Disclosed herein are non-covalently coupled proteins. Some aspects relate to a nucleic acid (e.g. an expression vector) encoding a protein, or encoding at least part of a protein. The proteins may include an endonuclease such as an RNA-guided endonuclease. A protein of the non-covalently coupled proteins may include a portion of an endonuclease. A protein of the non-covalently coupled proteins may include a portion of a ligase. The proteins may include a ligase such as a DNA ligase. A protein of the non-covalently coupled proteins may include a fusion protein.

[00105] The non-covalently coupled proteins may be bound together through heterodimerization domains. Examples of heterodimerization domains may include a leucine zipper, PDZ domain, streptavidin, streptavidin binding protein, foldon domain, hydrophobic moiety, or a functional binding fragment thereof. A heterodimerization domain may include a leucine zipper. A heterodimerization domain may include a PDZ domain. A heterodimerization domain may include a streptavidin. A heterodimerization domain may include a streptavidin binding protein. A heterodimerization domain may include a foldon domain. A heterodimerization domain may include a hydrophobic moiety. A heterodimerization domain may include an antibody or antibody fragment. The non-covalently coupled proteins may be bound together through inteins.

[00106] The endonuclease and ligase may be coupled together by a separate molecule. The separate molecule may comprise a nucleic acid (e.g. a guide nucleic acid). The ligase may include a hairpin binding motif, where the RNA-guided endonuclease and the DNA ligase are coupled with the nucleic acid. The nucleic acid may include a scaffold that binds the RNA-guided endonuclease and a hairpin that binds to the hairpin binding motif. The hairpin binding motif may include an MS2 coat protein (MCP) peptide. The hairpin may include an MS2 hairpin.

[00107] The endonuclease and ligase may be coupled together by a heterobifunctional molecule. The heterobifunctional molecule may include an endonuclease binding domain and a DNA ligase binding domain. The heterobifunctional molecule may include an endonuclease binding domain. The endonuclease binding domain may include a heterodimerization domain. The endonuclease binding domain may include an antibody or antibody binding fragment. The heterobifunctional molecule may include a ligase binding domain such as a DNA ligase binding domain. The DNA ligase binding domain may include a heterodimerization domain. The DNA ligase binding domain may include an antibody or antibody binding fragment. The heterobifunctional molecule may include a small molecule. The small molecule may comprise a proteolysis targeting chimera (PROTAC), or a related heterobifunctional molecule.

[00108] Some aspects include a protein complex, comprising: an RNA-guided endonuclease bound to a DNA ligase. The endonuclease and the DNA ligase may be bound together through heterodimerization domains. The protein complex of embodiment 75, wherein the heterodimerization domains may comprise leucine zippers, PDZ domains, streptavidin and streptavidin binding protein, foldon domains, hydrophobic polypeptides, an antibody that binds the Cas nickase, or an antibody that binds the DNA ligase, or one or more binding fragments thereof. The protein complex may be included in a cell. The cell may further include a heterologous RNA-guided endonuclease and a DNA ligase that was introduced into the cell. The cell may further include a nuclease that is different from the RNA-guided endonuclease.

Guide Nucleic Acids

[00109] Disclosed herein are guide nucleic acids. The guide nucleic acid may be included in a composition, system or method disclosed herein. Some aspects relate to a nucleic acid (e.g. DNA or an expression vector) that encodes a guide nucleic acid such as a guide RNA. Provided herein are guide nucleic acids (e.g., gRNAs) that direct a programmable endonuclease (e.g., a nCas9) to a target nucleic acid (e.g. a genomic locus). The guide nucleic acid may guide an RNA-guided endonuclease to a target nucleic acid locus for nucleic acid replacement or gene editing at the locus. A guide nucleic acid of the present disclosure may facilitate a donor strand to be inserted into a target site of the target nucleic acid. A guide nucleic acid of the present disclosure may facilitate editing of a nucleic acid sequence at a target site of the target nucleic acid. The guide nucleic acid may, in some instances, also act as a splint for a DNA ligase described herein, such as for ligating two nucleic acid strands base paired to a portion of the guide nucleic acid. The guide nucleic acid may be single stranded. The guide nucleic acid may include RNA. The guide nucleic acid may be RNA. The guide nucleic acid may include a guide RNA (gRNA). In some cases, a guide nucleic acid may include DNA.

[00110] The guide nucleic acid may be non-naturally occurring. The guide nucleic acid may be engineered. The guide nucleic acid may be synthetic. The guide nucleic acid may be pre-synthesized. The guide nucleic acid may be added to a subject or a cell. In some aspects, the guide nucleic acid does not include a template for a polymerase.

[00111] The guide nucleic acid may include an integrating nucleic acid binding site. The integrating nucleic acid binding site may be referred to as a “donor binding site.”

[00112] Disclosed herein are guide nucleic acids, comprising: a spacer reverse complementary to a first region of a target nucleic acid; a scaffold configured to bind to an endonuclease; and an integrating nucleic acid binding site and optionally a flap binding site reverse complementary to a nucleic acid flap.

[00113] In some aspects, the guide nucleic acid comprises a spacer complementary to a genomic locus in a cell; a scaffold for complexing with the at least one endonuclease; a donor binding site that is at least partially complementary to a donor strand; a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus; or a combination thereof. In some aspects, the guide nucleic acid can direct the at least one endonuclease to cleave at least one strand of the genomic locus. In some aspects, the guide nucleic acid can be at least partially complementary to the donor strand or at least partially complementary to a genomic flap (e.g., a genomic nucleic acid sequence that is displaced and become single-stranded when the guide nucleic acid recruits the endonuclease to the genomic locus). In some aspects, the guide nucleic acid, being at least partially complementary to the donor strand or at least partially complementary to a genomic flap, brings the donor strand to close proximity of the cleaving of the genomic locus.

[00114] Disclosed herein, in some embodiments, are guide nucleic acids comprising a scaffold. The scaffold may bind a nuclease. The scaffold may bind a Cas nuclease. The scaffold may bind a nickase. The scaffold may bind a Cas nickase. The scaffold may bind an *S. Pyogenes* Cas9 nuclease. The scaffold may bind an *S. Pyogenes* Cas9 nickase. The scaffold may include a scaffold nucleic acid sequence. A system described herein may include a first guide nucleic acid. The system can include a second guide nucleic acid. The first guide nucleic acid may bind to a first Cas nickase. The second guide nucleic acid may bind to a second Cas nickase.

[00115] A guide nucleic acid may include any aspect of (i)-(iv): (i) a spacer complementary to a region of a genomic locus of a genomic strand, (ii) a scaffold for complexing with an RNA-guided endonuclease, (iii) a donor binding site that is at least partially complementary to an integrating nucleic acid, or (iv) a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus. A guide nucleic acid may include any aspect of (i)-(iii): (i) a spacer complementary to a region of a genomic locus of a

genomic strand, (ii) a scaffold for complexing with an RNA-guided endonuclease, or (iii) a donor binding site that is at least partially complementary to a splinting nucleic acid. A component of (i), (ii), or (iii) may be included in a single guide nucleic acid, or may be split between or collectively included among multiple guide nucleic acids.

[00116] In some aspects, the guide nucleic acid comprises a modified internucleoside linkage. In some aspects, the modified internucleoside linkage comprises a phosphorothioate linkage. In some aspects, the modified internucleoside linkage is between any of the 4 terminal nucleosides at a 5' end or at a 3' end of the guide nucleic acid. The guide nucleic acid may include multiple modified internucleoside linkages. For example, the guide nucleic acid may include modified internucleoside linkages at nucleic acids of the 5' and 3' ends of the guide nucleic acid, such as between the last 4 nucleic acids at the 5' end and between the last 4 nucleic acids at the 3' end. In some aspects, the guide nucleic acid comprises a modified nucleoside. In some aspects, the modified nucleoside comprises a locked nucleic acid (LNA), a 2' fluoro, a 2' O-alkyl, or a combination thereof. The modified nucleoside may include an LNA, a 2' fluoro, a 2' O-alkyl, a methylated cytosine, an inverted thymidine, or a combination thereof. The modified nucleoside may include an LNA. The modified nucleoside may include a 2' fluoro. The modified nucleoside may include a 2' O-alkyl. The modified nucleoside may include a methylated cytosine. In some aspects, the modified nucleoside is any of the 3 terminal nucleosides at a 5' end or at a 3' end of the guide nucleic acid. The guide nucleic acid may include multiple modified nucleosides. For example, the guide nucleic acid may include modified nucleosides at nucleic acids of the 5' and 3' ends of the guide nucleic acid, such as the last 3 nucleic acids at the 5' end and the last 3 nucleic acids at the 3' end.

[00117] In some aspects, the guide nucleic acid comprises at least one nucleic acid modification. In some aspect, the at least nucleic acid modification comprises modifying a backbone, a sugar, a base, or a combination thereof of the guide nucleic acid. In some aspects, the at least one nucleic acid modification can increase resistance of the guide nucleic acid to degradation (e.g., against nuclease degradation or hydrolysis). In some aspects, the at least one nucleic acid modification can increase the complexing of the guide nucleic acid to the at least one endonuclease. In some aspects, the at least one nucleic acid modification can increase the complexing of the guide nucleic acid to the donor strand. In some aspects, the at least one nucleic acid modification can increase the complexing of the guide nucleic acid to the genomic locus via by being complementary to the genomic flap.

[00118] In some aspects, the guide nucleic acid comprises at least one, two, three, four, five, six, seven, eight, nine, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more

nucleic acid modifications. In some aspects, nucleic acid modification can occur at 3'OH group, 5'OH group, at the backbone, at the sugar component, or at the nucleotide base. Nucleic acid modification can include non-naturally occurring linker molecules of interstrand or intrastrand cross links. In one aspect, the modified nucleic acid comprises modification of one or more of the 3'OH or 5'OH group, the backbone, the sugar component, or the nucleotide base, or addition of non-naturally occurring linker molecules. In some aspects, modified backbone comprises a backbone other than a phosphodiester backbone. In some aspects, a modified sugar comprises a sugar other than deoxyribose (in modified DNA) or other than ribose (modified RNA). In some aspects, a modified base comprises a base other than adenine, guanine, cytosine, thymine or uracil. In some aspects, the guide nucleic acid comprises at least one modified base. In some instances, the guide nucleic acid comprises at least one, two, three, four, five, six, seven, eight, nine, 10, 15, 20, or more modified bases. In some cases, the nucleic acid modifications to the base moiety include natural and synthetic modifications of adenine, guanine, cytosine, thymine, or uracil, and purine or pyrimidine bases.

[00119] In some aspects, the at least one nucleic acid modification of the guide nucleic acid comprises a modification of any one of or any combination of: 2' modified nucleotide comprising 2'-O-methyl, 2'-O-methoxyethyl (2'-O-MOE), 2'-O-aminopropyl, 2'-deoxy, 2'-deoxy-2'-fluoro, 2'-O-aminopropyl (2'-O-AP), 2'-O-dimethylaminoethyl (2'-O-DMAOE), 2'-O-dimethylaminopropyl (2'-O-DMAP), 2'-O-dimethylaminoethoxyethyl (2'-O-DMAEOE), or 2'-O-N-methylacetamido (2'-O-NMA); modification of one or both of the non-linking phosphate oxygens in the phosphodiester backbone linkage; modification of one or more of the linking phosphate oxygens in the phosphodiester backbone linkage; modification of a constituent of the ribose sugar; replacement of the phosphate moiety with “dephospho” linkers; modification or replacement of a naturally occurring nucleobase; modification of the ribose-phosphate backbone; modification of 5' end of polynucleotide; modification of 3' end of polynucleotide; modification of the deoxyribose phosphate backbone; substitution of the phosphate group; modification of the ribophosphate backbone; modifications to the sugar of a nucleotide; modifications to the base of a nucleotide; or stereopure of nucleotide. Non limiting examples of nucleic acid modification to the guide nucleic acid can include: modification of one or both of non-linking or linking phosphate oxygens in the phosphodiester backbone linkage (e.g., sulfur (S), selenium (Se), BR3 (wherein R can be, e.g., hydrogen, alkyl, or aryl), C (e.g., an alkyl group, an aryl group, and the like), H, NR2, wherein R can be, e.g., hydrogen, alkyl, or aryl, or wherein R can be, e.g., alkyl or aryl); replacement of the phosphate moiety with “dephospho” linkers (e.g., replacement with methyl phosphonate, hydroxylamino, siloxane, carbonate, carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide,

thioformacetal, formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo, or methyleneoxymethylimino); modification or replacement of a naturally occurring nucleobase with nucleic acid analog; modification of deoxyribose-phosphate or ribose-phosphate backbone (e.g., modifying the ribose-phosphate backbone to incorporate phosphorothioate, phosphonothioacetate, phosphoroselenates, boranophosphates, borano phosphate esters, hydrogen phosphonates, phosphonocarboxylate, phosphoroamidates, alkyl or aryl phosphonates, phosphonoacetate, or phosphotriesters; modification of 5' end (e.g., 5' cap or modification of 5' cap -OH) or 3' end of the nucleic acid sequence (3' tail or modification of 3' end -OH); substitution of the phosphate group with methyl phosphonate, hydroxylamino, siloxane, carbonate, carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide, thioformacetal, formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo, or methyleneoxymethylimino; modification of the ribophosphate backbone to incorporate morpholino (phosphorodiamidate morpholino oligomer PMO), cyclobutyl, pyrrolidine, or peptide nucleic acid (PNA) nucleoside surrogates; modifications to the sugar of a nucleotide to incorporate locked nucleic acid (LNA), unlocked nucleic acid (UNA), ethylene nucleic acid (ENA), constrained ethyl (cEt) sugar, or bridged nucleic acid (BNA); modification of a constituent of the ribose sugar (e.g., 2'-O-methyl, 2'-O-methoxy-ethyl (2'-MOE), 2'-fluoro, 2'-aminoethyl, 2'-deoxy-2'-fuloarabinou-cleic acid, 2'-deoxy, 2'-O-methyl, 3'-phosphorothioate, 3'-phosphonoacetate (PACE), or 3'-phosphonothioacetate (thioPACE)); modification to the base of a nucleotide (of A, T, C, G, or U); and stereopure of nucleotide (e.g., S conformation of phosphorothioate or R conformation of phosphorothioate).

[00120] In some aspects, the nucleic acid modification comprises at least one substitution of one or both of non-linking phosphate oxygen atoms in a phosphodiester backbone linkage of the guide nucleic acid. In some aspects, the at least one nucleic acid modification of the guide nucleic acid comprises a substitution of one or more of linking phosphate oxygen atoms in a phosphodiester backbone linkage of the guide nucleic acid. A non-limiting example of a nucleic acid modification of a phosphate oxygen atom is a sulfur atom. In some aspects, the nucleic acid modification comprises at least one modification to a sugar. In some aspects, the nucleic acid modification comprises at least one nucleic acid modification to the sugar comprising a modification of a constituent of the sugar, where the sugar is a ribose sugar. In some aspects, the nucleic acid modification of the guide nucleic acid comprises at least one modification to the constituent of the ribose sugar of the nucleotide of the guide nucleic acid comprising a 2'-O-Methyl group. In some aspects, the nucleic acid modification comprises at least one modification comprising replacement of a phosphate moiety of the guide nucleic acid with a

dephospho linker. In some aspects, the nucleic acid modification of comprises at least one modification of a phosphate backbone. In some aspects, the modification comprises a phosphorothioate group. In some aspects, the nucleic acid modifications comprises at least one modification comprising a modification to a base of a nucleotide of the guide nucleic acid. In some aspects, the nucleic acid modifications comprises at least one modification comprising an unnatural base of a nucleotide. In some aspects, the nucleic acid modifications comprises at least one modification comprising at least one stereopure nucleic acid. In some aspects, the at least one nucleic acid modification can be positioned proximal to a 5' end of the guide nucleic acid. In some aspects, the at least one nucleic acid modification can be positioned proximal to a 3' end of the guide nucleic acid. In some aspects, the at least one nucleic acid modification can be positioned proximal to both 5' and 3' ends of the guide nucleic acid.

[00121] In some aspects, the guide nucleic acid described herein comprises a backbone comprising a plurality of sugar and phosphate moieties covalently linked together. In some cases, a backbone of the guide nucleic acid comprises a phosphodiester bond linkage between a first hydroxyl group in a phosphate group on a 5' carbon of a deoxyribose in DNA or ribose in RNA and a second hydroxyl group on a 3' carbon of a deoxyribose in DNA or ribose in RNA. In some aspects, a backbone of the guide nucleic acid can lack a 5' reducing hydroxyl, a 3' reducing hydroxyl, or both, capable of being exposed to a solvent. In some aspects, a backbone of the guide nucleic acid can lack a 5' reducing hydroxyl, a 3' reducing hydroxyl, or both, capable of being exposed to nucleases. In some aspects, a backbone of the guide nucleic acid can lack a 5' reducing hydroxyl, a 3' reducing hydroxyl, or both, capable of being exposed to hydrolytic enzymes. In some instances, a backbone of the guide nucleic acid can be represented as a polynucleotide sequence in a circular 2-dimensional format with one nucleotide after the other. In some instances, a backbone of the guide nucleic acid can be represented as a polynucleotide sequence in a looped 2-dimensional format with one nucleotide after the other. In some cases, a 5' hydroxyl, a 3' hydroxyl, or both, are joined through a phosphorus-oxygen bond. In some cases, a 5' hydroxyl, a 3' hydroxyl, or both, are modified into a phosphoester with a phosphorus-containing moiety. In some aspects, the guide nucleic acid comprises at least one nucleic acid modification comprising any one of: 5' adenylate, 5' guanosine-triphosphate cap, 5'N7-Methylguanosine-triphosphate cap, 5'triphosphate cap, 3'phosphate, 3'thiophosphate, 5'phosphate, 5'thiophosphate, Cis-Syn thymidine dimer, trimers, C12 spacer, C3 spacer, C6 spacer, dSpacer, PC spacer, rSpacer, Spacer 18, Spacer 9, 3'-3' modifications, 5'-5' modifications, abasic, acridine, azobenzene, biotin, biotin BB, biotin TEG, cholesteryl TEG, desthiobiotin TEG, DNP TEG, DNP-X, DOTA, dT-Biotin, dual biotin, PC biotin, psoralen C2, psoralen C6, TINA, 3'DABCYL, black hole quencher 1, black hole quencher 2, DABCYL SE,

dT-DABCYL, IRDye QC-1, QSY-21, QSY-35, QSY-7, QSY-9, carboxyl linker, thiol linkers, 2'deoxyribonucleoside analog purine, 2'deoxyribonucleoside analog pyrimidine, ribonucleoside analog, 2'-O-methyl ribonucleoside analog, sugar modified analogs, wobble/universal bases, fluorescent dye label, 2'fluoro RNA, 2'O-methyl RNA, methylphosphonate, phosphodiester DNA, phosphodiester RNA, phosphothioate DNA, phosphorothioate RNA, UNA, LNA, cEt, pseudouridine-5'-triphosphate, 5-methylcytidine-5'-triphosphate, 2-O-methyl -phosphorothioate or any combinations thereof.

[00122] A nucleic acid modification can also be a phosphorothioate substitute. In some cases, a natural phosphodiester bond can be susceptible to rapid degradation by cellular nucleases and; a modification of internucleotide linkage using phosphorothioate (PS) bond substitutes can be more stable towards hydrolysis by cellular degradation. A modification can increase stability in a polynucleic acid. A modification can also enhance biological activity. In some cases, a phosphorothioate enhanced RNA polynucleic acid can inhibit RNase A, RNase T1, calf serum nucleases, or any combinations thereof. These properties can allow the use of PS-RNA polynucleic acids to be used in applications where exposure to nucleases is of high probability *in vivo* or *in vitro*. For example, phosphorothioate (PS) bonds can be introduced between the last 3-5 nucleotides at the 5'-or 3'-end of a polynucleic acid which can inhibit exonuclease degradation. In some cases, phosphorothioate bonds can be added throughout an entire polynucleic acid to reduce attack by endonucleases. In some aspects, the guide nucleic acid comprises at least one, two, three, four, five, six, seven, eight, nine, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 50, 100, or more internucleotide linkage comprising PS bond. In some aspects, the guide nucleic acid comprises only PS bond as the internucleotide linkage modification. In some aspects, all internucleotide linkages of the guide nucleic acid herein are fully PS-modified or include phosphorothioate internucleotide linkages.

[00123] The guide nucleic acid may include a hairpin. The hairpin may bind to a hairpin binding motif such as a hairpin binding motif on a DNA ligase. The hairpin may include an MS2 hairpin A hairpin such as an MS2 hairpin may be useful for recruiting a DNA ligase that includes an MCP peptide.

[00124] The guide nucleic acid may include any aspect included in **Fig. 1A-6C**. **Table 8** illustrates non-limiting examples of some of the guide nucleic acids described herein. Some of the guide nucleic acids in the table include nucleic acid modifications.

Table 8. Examples of nucleic acid sequences

Name	Nucleic acid sequence (5' to 3')	SEQ ID NO:
BFP	atggtgagcaagggcgaggagctgttcaccggggtggtgccatcctggtcgagctggacggcgac gtaaaccggccacaagttcagcgtgtccggcgaggggcgaggcgatgccacctacggcaagctgac cctgaagttcatctgcaccaccggcaagctgccGgtgccctggcccaccCTCgtgaccacctga ccCATggcgtgcagtgttcagccgctaccccgaccacatgaagcagcagcagcttctcaagtccg ccatgcccgaaggctacgtccaggagcgaccatcttctcaaggacgacggcaactacaagacc ggcccgagggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttca aggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatcat ggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggca cggtgcagctcgccgaccactaccagcagaacacccccatggcgacggccccgtgctgctgcc gacaaccactacgtgagcaccagtcgaagctgagcaaaagacccaacgagaagcgcgatcacat ggtcctgctggagttcgtgaccggcgccggatcactctcgcatggacgagctgtacaagTAA	97
Rep1.BFP.Fwd Guide	mC*mU*mG*AAGUUCaucugcaccacguuuAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCGUGGGCCAG GGCACCGGCAGCUUGCCGGUGGUGCAGAUgMA*mA*mC*U	98
Rep1.BFP.Rev Guide	mG*mA*mC*GUAGCCUUCGGGCAUGGGUUUAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCUGAAGCAGC ACGACUUCUUAAGUCCGCCAUGCCCCGAAGmG*mC*mU*A	99
Rep1.BFP.Fwd Guide.SpPAM mut	mC*mU*mG*AAGUUCaucugcaccacguuuAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCGUGGGCCAG GGCACCGGCAGCUUGCCGGUUGUGCAGAUgMA*mA*mC*U	100
Rep1.BFP.Rev Guide.SpPAM mut	mG*mA*mC*GUAGCCUUCGGGCAUGGGUUUAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCUGAAGCAGC ACGACUUCUUAAGUCCGCCAUGCCCCGAAGmG*mC*mU*A	101
Rep1.BFP2GF P.TopDonor.5P	/5Phos/caccggcaagctgccGgtgccctggcccaccCTCgtgaccacctgaccTACggc gtgcagtgtctcagccgtaccccgaccaca	102
Rep1.BFP2GF P.BotDonor.5P	/5Phos/tggcggactgaagaagtcgtgctgcttcatgtggtcgggtagcggctgaagcactgcac gccGTAggtcagggtggtcacGAGg	103
Rep1.BFP2GF P.TopDonor.Recoded.5P	/5Phos/caccggcaagctgccGgtgccctggcccacTCTTGTGACCACCTTGAC CtACGGTGTCCAGTGTTTTAGCAGGTATCCGGATCACA	104
Rep1.BFP2GF P.BotDonor.Recoded.5P	/5Phos/tggcggactgaagaagtcgtgctgcttcaTGTGATCCGGATACCTGCTA AAACACTGGACACCGTaGGTCAAGGTGGTCACAAGA	105
Rep1.BFP2GF P.TopDonor.SpPAMmut.Recoded.5P	/5Phos/AaccggcaagctgccGgtgccctggcccacTCTTGTGACCACCTTGAC CtACGGTGTCCAGTGTTTTAGCAGGTATCCGGATCACA	106
Rep1.BFP2GF P.BotDonor.SpPAMmut.Recoded.5P	/5Phos/tggcTgactgaagaagtcgtgctgcttcaTGTGATCCGGATACCTGCT AAAACACTGGACACCGTaGGTCAAGGTGGTCACAAGA	107

Name	Nucleic acid sequence (5' to 3')	SEQ ID NO:
Rep1.BFP2GF P.TopDonor.R ecoded.5P.end Phos	/5Phos/c*a*c*cggcaagctgccGgtgccctggccacTCTTGTGACCACCTTG ACcTACGGTGTCCAGTGTtTTAGCAGGTATCCGGATC*A*C*A	108
Rep1.BFP2GF P.BotDonor.Re coded.5P.endP hos	/5Phos/t*g*g*cggactgaagaagtcgtgctgcttcaTGTGATCCGGATACCTG CTAAACACTGGACACCGTaGGTCAAGGTGGTCACA*A*G*A	109
Rep2.BFP.Fwd Guide	mC*mU*mG*AAGUUCAUCUGCACCACGUUUAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCCUACGGCAAG CUGACCMC*mU*mG*A	110
Rep2.BFP.Rev Guide	mG*mA*mC*GUAGCCUUCGGGCAUGGGUUUAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCAGAUGGUGC GCUCCUGmG*mA*mC*G	111
Rep2.BFP2GF P.TopDonor.S pPAMmut.5P	/5Phos/AaccggcaagctgccGgtgccctggccaccCTCgtgaccaccctgaccTACgg cgtgcagtgttcagccgctaccccgaccacatgaagcagcagcacttctcaagtcAgccatgcc gaaggctacgtccaggagcgcaccatct	112
Rep2.BFP2GF P.BotDonor.Sp PAMmut.5P	/5Phos/tggcTgacttgaagaagtcgtgctgcttcatgtggtcgggtagcggctgaagcactgca cgccGTAggtcagggtggtcacGAGggtgggcccagggcacCggcagcttgccggtTgtgca gatgaacttcagggtcagcttgccgtag	113
Rep2.BFP2GF P.TopDonor.S pPAMmut.Rec oded.5P	/5Phos/AACCGGTAAGTTGCCAGTCCCGTGGCCTACTCTTGTGA CCACCTTGACcTACGGTGTCCAGTGTtTTAGCAGGTATCCGGA TCACATGAAACAGCATGACTTCTTTAAATCAGCTATGcccgaagg ctacgtccaggagcgcaccatct	114
Rep2.BFP2GF P.BotDonor.Sp PAMmut.Reco ded.5P	/5Phos/TAGCTGATTTAAAGAAGTCATGCTGTTTCATGTGATCC GGATACCTGCTAAAACACTGGACACCGTaGGTCAAGGTGGTC ACAAGAGTAGGCCACGGGACTGGCAACTTACCGGTTgtgcagatg aacttcagggtcagcttgccgtag	115
Rep2.BFP2GF P.TopDonor.5 P	/5Phos/caccggcaagctgccGgtgccctggccaccCTCgtgaccaccctgaccTACggc gtgcagtgttcagccgctaccccgaccacatgaagcagcagcacttctcaagtcgcccagatgcccga aggctacgtccaggagcgcaccatct	116
Rep2.BFP2GF P.BotDonor.5P	/5Phos/tggcggacttgaagaagtcgtgctgcttcatgtggtcgggtagcggctgaagcactgcac gccGTAggtcagggtggtcacGAGggtgggcccagggcacCggcagcttgccggtggtgcag atgaacttcagggtcagcttgccgtag	117
Rep2.BFP2GF P.TopDonor.S pPAMmut.Rec oded.5P.endPh os	/5Phos/A*A*A*C*CGGTAAGTTGCCAGTCCCGTGGCCTACTCTTGT GACCACCTTGACcTACGGTGTCCAGTGTtTTAGCAGGTATCCG GATCACATGAAACAGCATGACTTCTTTAAATCAGCTATGcccga aggctacgtccaggagcgcacca*t*c*t	118
Rep2.BFP2GF P.BotDonor.Sp PAMmut.Reco ded.5P.endPho s	/5Phos/T*A*G*CTGATTTAAAGAAGTCATGCTGTTTCATGTGAT CCGGATACCTGCTAAAACACTGGACACCGTaGGTCAAGGTGG TCACAAGAGTAGGCCACGGGACTGGCAACTTACCGGTTgtgcag atgaacttcagggtcagcttgccg*t*a*g	119
Rep2.CBX1.F wdGuide	mG*mA*mA*AGCUGGCGGGCACUAUGGUUUAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA	120

Name	Nucleic acid sequence (5' to 3')	SEQ ID NO:
	UCAACUUGAAAAAGUGGCACCGAGUCGGUGCGUCACCCUUU ACACCAmG*mA*mA*A	
Rep2.CBX1.R evGuide	mC*mU*mU*UGCCCUUUAACACUCGAGUUUAAGAGCUAUGC UGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUAGGAGG UACUCCAmC*mU*mU*U	121
Rep2.mGL- CBX1.TopDon or.SpPAMmut. 5P	ATGgtgagcaagggcgaggagctgttcacgggggtggtgcccatcctggtcgagctggacggcg acgtaaacggccacaagttcagcgtccgcgaggcgaggcgatgccaccaacggcaagct gacctgaagttcatctgcaccacggcaagctgccgtgccctggccaccctcgtgaccaccttag gctacggcggtggcctgttcgcccgtaccccgaccacatgaagcagcagcactcttcaagtccgc catgccgaaggctacgtccaggagcgcaccatctcttcaaggacgacggtacctacaagaccgc gccgaggtgaagttcgaggcgacacccctggtgaaccgcatcgtgctgaaggcgatcgacttcaag gaggacggcaacatcctggggcacaagctggagtacaacttaacagccacaaggtctatatcacg gccgacaagcagaagaacggcatcaaggtaacttaagaccgccacaacgttgaggacggcg cgtgcagctcggcaccactaccagcagaacacccccatcgcgacggccccgtgctgctgccg acaaccactacgtgaccatcagtcacaaactgagcaagaccccaacgagaagcgcgatcacatgg tctgaaggagaggggtgaccgccgggattacacatgacatggacgagctgtacaagtctggag gatctagcggaggatccGGGAAGAAACAAAACAAGAAGAAAGTGGAG GAGGTGCTAGAAGAGGAGGAAGAGGAATATGTGGTGGAAAA AGTTCTCGAtCGTCGAGTGGTAAAGGGCAAAGTGGAGTACCT CCTAAA	122
Rep2.mGL- CBX1.BotDon or.SpPAMmut. 5P	CGACGaTCGAGAActTTTTCCACCACATATTCCTCTTCCTCCTC TTCTAGCACCTCCTCCACTTTCTTCTTGTGTTTTGTTCTTCCCggat cctcgcgtagatcctccagactgtacagctcgtccatgtcatgttaatcccgcgcggtcaccctct ccttcaggaccatgtgatcgcttctcgttgggtcttctcagtttgactgatggctcaggtagtg ttgtcgggcagcagcagcgggcggtcgccgatgggggtgttctgctggtagtgtcgcgagctgc acggcgcgctcctcaacgttggcggtcttgaagttagccttgatgccgttctctgcttgcggcgt gatatagaccttgggtgttgaagtgtactccagcttgtgcccgaggtgttgcgctcctcttgaagt cgatgcccttcagcagcatcggttaccagggtgtcgccctcgaaacttcacctcgcgcggtgtt taggtaccgtcgtccttgaagagatgggtcgctcctggacgtagccttcgggcagtgccgactgaa gaagtctgctgcttcatgtggtcggggtagcgggcgaagcaggccagcgctagcctaagggtgt cacgaggggtgggcccagggcacgggcagcttgcgggtgtgcagatgaacttcagggtcagcttgc gttgggtgcatcgccctcgccctcgccgacgctgaacttggccgtttacgtcgccgtccagct cgaccaggatgggcaccaccccggtgaacagctcctcgcccttgcctacCATAGTGCCCCG CCAGCTTTCTGGTGTAAGGGGTGAC	123
CBX1-001 Exon 2 (includes beginning of ORF)	CAGCGTCACCCTTTACACCAGAAAGCTGGCGGGCACTATGGG GAAAAAACAAAACAAGAAGAAAGTGGAGGAGGTGCTAGAA GAGGAGGAAGAGGAATATGTGGTGGAAAAAGTTCTCGACCG TCGAGTGGTAAAGGGCAAAGTGGAGTACCTCCTAAAGTGA AGGGATTCTCAGA	124

[00125] The guide nucleic acid may include a sequence of linking nucleic acids (e.g. linking RNA or DNA nucleotides) between components of the guide nucleic acid. For example, the guide nucleic acid may include a sequence of linking nucleic acids between any of the following components: a spacer, a scaffold, a donor binding site, or a flap binding site. The guide nucleic acid may include a sequence of linking nucleic acids between a spacer, a scaffold,

or a donor binding site. The guide nucleic acid include a sequence of linking nucleic acids between the scaffold and the donor binding site. The guide nucleic acid may include a sequence of linking nucleic acids between a spacer and a scaffold. The guide nucleic acid may include multiple sequences of linking nucleic acids between components.

[00126] The sequence of linking nucleic acids may include any base, such as A, U, T, G, or C, or a combination thereof. The sequence of linking nucleic acids may include A, T, G, or C, or a combination thereof. The sequence of linking nucleic acids may include A, U, G, or C, or a combination thereof. The sequence of linking nucleic acids may include a series of As. The sequence of linking nucleic acids may include a series of Ts. The sequence of linking nucleic acids may include a series of Us. The sequence of linking nucleic acids may include a series of Cs. The sequence of linking nucleic acids may include a series of Gs.

[00127] The sequence of linking nucleic acids may include a length, such as a number of nucleotides. The length may include 1, 2, 3, 4, 5, 6, 7, 8, 9 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides, or a range defined by any two of the aforementioned numbers of nucleotides. The length may include at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 nucleotides. In some aspects, the length may be less than 2, less than 3, less than 4, less than 5, less than 6, less than 7, less than 8, less than 9 10, less than 11, less than 12, less than 13, less than 14, less than 15, less than 16, less than 17, less than 18, less than 19, less than 20, less than 21, less than 22, less than 23, less than 24, less than 25, less than 30, less than 35, less than 40, less than 45, less than 50, less than 55, less than 60, less than 65, less than 70, less than 75, less than 80, less than 85, less than 90, less than 95, or less than 100 nucleotides.

[00128] Some aspects relate to a guide nucleic acid comprising: a spacer that is at least partially complementary to a genomic locus in a cell; a scaffold for complexing with a RNA-guided endonuclease; and a donor binding site that is at least partially complementary to an integrating nucleic acid. The guide nucleic acid may further comprise a flap binding site that is at least partially complementary to a genomic sequence of the genomic locus. The guide nucleic acid may further comprise at least one nucleic acid modification. The at least one nucleic acid modification may comprise a modification to a backbone, a sugar, a base, or a combination thereof. The guide nucleic acid may comprise RNA.

[00129] Some aspects include a guide nucleic acid, comprising: a spacer at least partially reverse complementary to a first region of a target nucleic acid; a scaffold configured to bind to an endonuclease; and a flap binding site at least partially reverse complementary to a nucleic acid flap, and an integrating nucleic acid binding site.

Integrating nucleic acids

[00130] Disclosed herein are integrating nucleic acids. The integrating nucleic acid may be included in a composition, system, or method disclosed herein. Some aspects relate to a nucleic acid that encodes an integrating nucleic acid. Provided herein are integrating nucleic acids that are inserted into a target nucleic acid such as a host genome at a genetic locus. For example, the integrating nucleic acid may replace a nucleic acid in the target nucleic acid. The integrating nucleic acid may be referred to as a “donor nucleic acid,” “donor” or “donor strand.” Where a genomic locus is described, a genetic locus may be included, or vice versa. For example, the locus may be part of a host genome or may be a part of a non-genome nucleic acid. The donor may include DNA. Likewise, the target nucleic acid may include DNA. In some cases, the donor may include RNA, for example when a target nucleic acid includes RNA. The integrating nucleic acid may include any insert, such as a gene or a regulatory element, to be inserted at a genomic locus of a target nucleic acid. The donor strand may include a sequence that is at least partially homologous to the genomic locus. The integrating nucleic acid may, in some instances, also act as a splint for a DNA ligase described herein, such as for ligating two nucleic acid strands base paired to a portion of the splinting integrating nucleic acid. In some cases, the splint includes one strand of the integrating nucleic acid, and the portion being ligated may be another strand of the integrating nucleic acid. In some cases, the splint includes a strand of the integrating nucleic acid, and the portion being ligated may be an upstream or downstream portion of the same strand of the integrating nucleic acid. The integrating nucleic acid may be single stranded. The integrating nucleic acid may be double stranded. The integrating nucleic acid may be delivered as two strands. The integrating nucleic acid may be delivered as multiple strands, e.g. 2 strands.

[00131] The integrating nucleic acid may be non-naturally occurring. The integrating nucleic acid may be engineered. The integrating nucleic acid may be synthetic. The integrating nucleic acid may be pre-synthesized. The integrating nucleic acid may be added to a subject or a cell. In some aspects, the integrating nucleic acid does not include a template for a polymerase.

[00132] Disclosed herein are integrating nucleic acids, comprising: a double-stranded DNA region to be inserted into a target nucleic acid, wherein the double-stranded DNA region is flanked by at least one overhang comprising a flap binding site and/or guide binding site.

[00133] The integrating nucleic acid may be ligated into a target nucleic acid such as a genomic strand. The integrating nucleic acid may include a 5' end that may be ligated to a 3' terminus of a genomic strand generated by an RNA-guided endonuclease.

[00134] The donor may include any aspect included in **Fig. 1A-6C**. For example, the donor may include an aspect such as a guide binding site, a flap binding site, or an overhang. The donor may include a guide binding site. The donor may include 2 guide binding sites. The donor may include a flap binding site. The donor may include 2 flap binding sites. The donor may include an overhang. The donor may include 2 overhangs. The aspects may be included at a 5' end or a 3' end of the donor, or at both ends. A guide binding site or a flap binding site may be in an internal region of the donor.

[00135] Some aspects include an integrating nucleic acid, comprising: a double-stranded DNA region to be inserted into a target nucleic acid, wherein the double-stranded DNA region is flanked by at least one overhang comprising a flap binding site or guide binding site.

[00136] In some aspects, the integrating nucleic acid comprises a modified internucleoside linkage. In some aspects, the modified internucleoside linkage comprises a phosphorothioate linkage. In some aspects, the modified internucleoside linkage is between any of the 4 terminal nucleosides at a 5' end or at a 3' end of the integrating nucleic acid. The integrating nucleic acid may include multiple modified internucleoside linkages. For example, the integrating nucleic acid may include modified internucleoside linkages at nucleic acids of the 5' and 3' ends of the integrating nucleic acid, such as between the last 4 nucleic acids at the 5' end and between the last 4 nucleic acids at the 3' end. In some aspects, the integrating nucleic acid comprises a modified nucleoside. In some aspects, the modified nucleoside comprises a locked nucleic acid (LNA), a 2' fluoro, a 2' O-alkyl, a 5' O-methyl, a 2'-O-methyl, or a combination thereof. The modified nucleoside may include an LNA, a 2' fluoro, a 2' O-alkyl, a methylated cytosine, an inverted thymidine, or a combination thereof. The modified nucleoside may include an LNA. The modified nucleoside may include a 2' fluoro. The modified nucleoside may include a 2' O-alkyl. The modified nucleoside may include a methylated cytosine. In some aspects, the modified nucleoside is any of the 3 terminal nucleosides at a 5' end or at a 3' end of the integrating nucleic acid. The integrating nucleic acid may include multiple modified nucleosides. For example, the integrating nucleic acid may include modified nucleosides at nucleic acids of the 5' and 3' ends of the integrating nucleic acid, such as the last 3 nucleic acids at the 5' end and the last 3 nucleic acids at the 3' end. The integrating nucleic acid may include any modification such as a modified nucleoside or modified internucleoside linkage described in relation to guide nucleic acids, insofar as it does not interfere with the function of the integrating nucleic acid after it is ligated into a target nucleic acid such as a host genome.

The integrating nucleic acid may include any number or combination of modifications such as a number or combination described in relation to guide nucleic acids, insofar as it does not interfere with a function of the integrating nucleic acid. **Table 8** includes some examples of integrating nucleic acid sequences.

[00137] The integrating nucleic acid may include a methylated nucleotide. The integrating nucleic acid may include an unmethylated nucleotide. An example of a methylated nucleotide may include a nucleotide including methylated cytosine. The cytosine may be methylated at a C-5 position of the cytosine ring. An example of an unmethylated nucleotide may include an unmethylated cytosine. The unmethylated nucleotide may include a cytosine that is not methylated at a C-5 position of the cytosine ring.

Target nucleic acids

[00138] Disclosed herein are target nucleic acids. The target nucleic acid may include DNA. The target nucleic acid may be DNA. The target nucleic acid may include RNA. The target nucleic acid may be in a cell. The target nucleic acid may be methylated. The target nucleic acid may be unmethylated. The target nucleic acid may comprise a genome. The target nucleic acid may comprise genomic DNA. The target nucleic acid may comprise a chromosome. The target nucleic acid may comprise a gene.

[00139] The target nucleic acid may be in a subject. The target nucleic acid may be in a cell. The target nucleic acid may be in a test tube.

[00140] The target nucleic acid may be edited. The target nucleic acid may be edited *in vitro*. The target nucleic acid may be edited *in vivo*.

Systems

[00141] Described herein are systems for nucleic acid editing (also known as gene editing). The editing system may include an endonuclease such as an RNA-guided endonuclease, a guide nucleic acid, and an integrating nucleic acid. Where gene editing is described, it is contemplated that the editing may be of a gene, regulatory element, or any sequence of a nucleic acid. Also, where genome editing is described, such as genome editing at a genetic locus, it is contemplated that nucleic acid editing not comprising a genome may also be performed. For example, genome editing may refer to editing of a genome of an organism, or may include editing of a nucleic acid that is not part of a genome. The systems described herein may be used in gene editing methods.

[00142] Described herein, in some aspects, is a system comprising at least one endonuclease; at least one guide nucleic acid; at least one ligase; at least one donor strand; or a combination thereof. In some aspects, the guide nucleic acid directs the endonuclease to the genomic locus for cleaving at least one strand of the genomic locus, where, after cleavage, the donor strand is

ligated and thus incorporated into the genomic locus by the ligase. In some aspects, the system comprises: a first endonuclease to be complexed with a first guide nucleic acid, where the first endonuclease can be operatively coupled to a first ligase; and a second endonuclease to be complexed with a second guide nucleic acid, where the second endonuclease can be operatively coupled to a second ligase. In such system each of the first endonuclease and the second endonuclease can each cleave at least one strand of the genomic locus for incorporation of the donor strand.

[00143] In some aspects, the system comprises one, two, three, or more endonucleases. In some aspects, the system comprises one endonucleases. In some aspects, the two endonucleases can each be complexed with a different guide nucleic acid. In some aspects, the two endonucleases can each be operatively coupled to a ligase. In some aspects, the endonuclease is a programmable endonuclease. In some aspects, the endonuclease comprises a RNA-guided endonuclease, where the guide nucleic acid comprises a guide RNA. In some aspects, the endonuclease comprises a nickase, where the endonuclease only cleaves one strand (as opposed to making a double-stranded break). In some aspects, the endonuclease comprises a localization signal sequence to increase the accumulation of the endonuclease in the proximity of the genomic locus (e.g., in the nucleus). In some aspects, the endonuclease comprises at least one additional domain. In some aspects, the at least one additional domain is a dimerization domain. In some aspects, the endonuclease comprising a dimerization domain can be dimerized with a ligase to form a heterodimer. In some aspects, the at least one additional domain is a functional domain. For example, the functional domain can comprises a chromatin modifying domain or a cell penetrating peptide. In some aspects, the endonuclease comprises a linker, where the linker can covalently connect the endonuclease with another polypeptide (e.g., the ligase). In some aspects, the linker covalently connects the endonuclease to the at least one additional domain. In some aspects, the endonuclease comprises a tag, where the tag can be used for increasing expression, identifying, or purifying the endonuclease.

[00144] In some aspects, the system comprises one, two, three, or more guide nucleic acids. In some aspects, the system comprises one guide nucleic acid, where the one guide nucleic acid can be complexed with at least one endonuclease. In some aspects, the system comprises two guide nucleic acids, where the two guide nucleic acids can each be complexed with the at least one endonuclease. In some aspects, the guide nucleic acid comprises a spacer complementary to a genomic locus in a cell; a scaffold for complexing with the at least one endonuclease; a donor binding site that is at least partially complementary to a donor strand; a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus; or a combination thereof. In some aspects, the guide nucleic acid can direct the at least

one endonuclease to cleave at least one strand of the genomic locus. In some aspects, the guide nucleic acid can be at least partially complementary to the donor strand or at least partially complementary to a genomic flap (e.g., a genomic nucleic acid sequence that is displaced and become single-stranded when the guide nucleic acid recruits the endonuclease to the genomic locus). In some aspects, the guide nucleic acid, being at least partially complementary to the donor strand or at least partially complementary to a genomic flap, brings the donor strand to close proximity of the cleaving of the genomic locus. In some aspects, the guide nucleic acid comprises at least one nucleic acid modification. In some aspect, the at least nucleic acid modification comprises modifying a backbone, a sugar, a base, or a combination thereof of the guide nucleic acid. In some aspects, the at least one nucleic acid modification can increase resistance of the guide nucleic acid to degradation (e.g., against nuclease degradation or hydrolysis). In some aspects, the at least one nucleic acid modification can increase the complexing of the guide nucleic acid to the at least one endonuclease. In some aspects, the at least one nucleic acid modification can increase the complexing of the guide nucleic acid to the donor strand. In some aspects, the at least one nucleic acid modification can increase the complexing of the guide nucleic acid to the genomic locus via by being complementary to the genomic flap.

[00145] In some aspects, the system comprises one, two, three, or more ligase. In some aspects, the system comprises one ligase. In some aspects, the one ligase is operatively coupled with at least one endonuclease, where the ligase can ligate at least one end of the donor strand to the cleaved genomic locus, thus incorporating the donor strand into the genomic locus. In some aspects, the system comprises two ligases. In some aspects, the two ligases can each be operatively coupled to a different endonuclease, where the genomic locus is cleaved at two or more locations. In such scenario, the two ligases can each ligate one end of the donor strand to the cleaved genomic locus, thus incorporating the donor strand into the genomic locus. In some aspects, the ligase comprises a ligase that can ligate a substrate comprising DNA. In some aspects, the ligase comprises a ligase that can ligate a substrate comprising a DNA splint. In some aspects, the ligase comprises a ligase that can ligate a substrate comprising a DNA/RNA. In some aspects, the ligase comprises a ligase that can ligate a substrate comprising a RNA splint. In some aspects, the ligase comprises at least one additional domain. In some aspects, the at least one additional domain is a dimerization domain. In some aspects, the ligase comprising a dimerization domain can be dimerized with an endonuclease to form a heterodimer. In some aspects, the at least one additional domain is a functional domain. For example, the functional domain can comprises a chromatin modifying domain or a cell penetrating peptide. In some aspects, the ligase comprises a linker, where the linker can covalently connect the ligase with

another polypeptide (e.g., the endonuclease). In some aspects, the linker covalently connects the ligase to the at least one additional domain. In some aspects, the ligase comprises a tag, where the tag can be used for increasing expression, identifying, or purifying the ligase.

[00146] Disclosed herein are fusion proteins comprising: an RNA-guided endonuclease fused to a ligase. **Table 9** illustrates non-limiting examples of polypeptide and nucleic acid sequences encoding a fusion polypeptide comprising components (e.g., a endonuclease fused to a ligase) of a system described herein. SEQ ID NO: 125 illustrates a nucleic acid sequence encoding the polypeptide sequence of SEQ ID NO: 126, where SEQ ID NO: 126 illustrates a fusion protein (NLS-nCas9-linker-hLIG1(119-919)-bpNLS) comprising a N-terminus NLS followed by a endonuclease (nCas9) covalently connected to a ligase (hLIG1, 119-919 fragment) via a linker followed by a C-terminus NLS. SEQ ID NO: 127 illustrates a nucleic acid sequence encoding the polypeptide sequence of SEQ ID NO: 128, where SEQ ID NO: 128 illustrates a fusion protein (NLS-nCas9-linker-hLIG1(233-919)-bpNLS) comprising a N-terminus NLS followed by a endonuclease (nCas9) covalently connected to a ligase (hLIG1, 233-919 fragment) via a linker followed by a C-terminus NLS. SEQ ID NO: 129 illustrates a nucleic acid sequence encoding the polypeptide sequence of SEQ ID NO: 130, where SEQ ID NO: 130 illustrates a fusion protein (NLS-nCas9-linker-SplintR-bpNLS) comprising a N-terminus NLS followed by a endonuclease (nCas9) covalently connected to a ligase (SplintR) via a linker followed by a C-terminus NLS. SEQ ID NO: 131 illustrates a nucleic acid sequence encoding the polypeptide sequence of SEQ ID NO: 132, where SEQ ID NO: 132 illustrates a fusion protein (NLS-nCas9-linker-T4LIG-bpNLS) comprising a N-terminus NLS followed by a endonuclease (nCas9) covalently connected to a ligase (T4LIG) via a linker followed by a C-terminus NLS. SEQ ID NO: 133 illustrates a nucleic acid sequence encoding a endonuclease (nCas9) comprising a N-terminus NLS and a leucine zipper (LZ) dimerization domain. SEQ ID NO: 134 illustrates a fusion protein (NLS1-hFEN1-linker1-nCas9-linker2-T4LIG-NLS2) comprising first NLS (NLS1) at N-terminus followed by a exonuclease (hFEN1) covalently connected to a endonuclease (nCas9) via linker1 and further covalently connected to a ligase (T4LIG) via linker 2 followed by a second NLS (NLS2) at C-terminus. SEQ ID NO: 135 illustrates a fusion protein (NLS1-hFEN1-linker1-T4LIG-linker2-nCas9-NLS2) comprising a N-terminus NLS1 followed by a exonuclease (hFEN1) covalently connected to a ligase (T4LIG) via linker 1 and further covalently connected to a endonuclease (nCas9) via linker 2 followed by a C-terminus NLS2. SEQ ID NO: 136 illustrates a fusion protein (NLS1-nCas9-linker1-hFEN1-linker2-T4LIG-NLS2) comprising a N-terminus NLS1 followed by a endonuclease (nCas9) covalently connected to a exonuclease (hFEN1) via linker 1 and further covalently connected to a ligase (T4LIG) via linker 2 followed by a C-terminus NLS2. SEQ ID NO: 137 illustrates a fusion

protein (NLS1-T4LIG-linker1-nCas9-linker2-hFEN1-NLS2) comprising a N-terminus NLS1 followed by a ligase (T4LIG) covalently connected to an endonuclease (nCas9) via linker 1 and further covalently connected to an exonuclease (hFEN1) via linker 2 followed by a C-terminus NLS2. SEQ ID NO: 138 illustrates a fusion protein (NLS1-nCas9-linker1-T4LIG-linker2-hFEN1-NLS2) comprising a N-terminus NLS1 followed by an endonuclease (nCas9) covalently connected to a ligase (T4LIG) via linker 1 and further covalently connected to an exonuclease (hFEN1) via linker 2 followed by a C-terminus NLS2. SEQ ID NO: 139 illustrates a fusion protein (NLS1-T4LIG-linker1-hFEN1-linker2-nCas9-NLS2) comprising a N-terminus NLS1 followed by a ligase (T4LIG) covalently connected to an exonuclease (hFEN1) via linker 1 and further covalently connected to an endonuclease (nCas9) via linker 2 followed by a C-terminus NLS2. SEQ ID NO: 140 illustrates a fusion protein (NLS1-T5 EXO-linker1-nCas9-linker2-T4LIG-NLS2) comprising a N-terminus NLS1 followed by an exonuclease (EXO) covalently connected to an endonuclease (nCas9) via linker 1 and further covalently connected to a ligase (T4LIG) via linker 2 followed by a C-terminus NLS2. SEQ ID NO: 141 illustrates a nucleic acid sequence encoding a fusion protein (LZ-SplintR-bpNLS) comprising a ligase (SplintR) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 142 illustrates a nucleic acid sequence encoding a fusion protein (LZ-T4LIG-bpNLS) comprising a ligase (T4LIG) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 143 illustrates a nucleic acid sequence encoding a fusion protein (LZ-hLIG 233-919 polypeptide fragment-bpNLS) comprising a ligase (hLIG) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 144 illustrates a nucleic acid sequence encoding a fusion protein (LZ-hLIG1 119-919 polypeptide fragment-bpNLS) comprising a ligase (hLIG) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 145 illustrates a nucleic acid sequence encoding a fusion protein (T4-LZ) comprising a ligase (T4) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 146 illustrates a nucleic acid sequence encoding a fusion protein (LZ-hLIG4(1-620)) comprising a ligase polypeptide fragment (hLIG4(1-620)) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 147 illustrates a nucleic acid sequence encoding a fusion protein (LZ-nCas9) comprising an endonuclease (nCas9) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 148 illustrates a nucleic acid sequence encoding a fusion protein (SplintR-LZ) comprising a ligase (SplintR) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 149 illustrates a nucleic acid sequence encoding a fusion protein (hLIG4(1-620)-LZ) comprising a ligase polypeptide fragment (hLIG4(1-620)) fused to a dimerization domain (LZ) and a NLS. SEQ ID NO: 150 illustrates a nucleic acid sequence encoding a fusion protein (nCas9-hLIG4(1-620)) comprising a ligase polypeptide fragment (hLIG4(1-620)) fused to an endonuclease (nCas9) and a NLS. SEQ ID NO: 151 illustrates a nucleic acid sequence encoding a fusion protein (T4-nCas9)

comprising a ligase (T4) fused to an endonuclease (nCas9) and a NLS. SEQ ID NO: 152 illustrates a nucleic acid sequence encoding a fusion protein (SplintR-nCas9) comprising a ligase (SplintR) fused to an endonuclease (nCas9) and a NLS. SEQ ID NO: 153 illustrates a nucleic acid sequence encoding a fusion protein (hLIG4(1-620)-nCas9) comprising a ligase polypeptide fragment (hLIG4(1-620)) fused to an endonuclease (nCas9) and a NLS.

Table 9. Non-limiting examples of fusion protein polypeptide sequence or nucleic acid sequence encoding the fusion protein

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
NLS-nCas9-linker-hLIG1(19-919)-bpNLS	atgaaacggacagccgacggaagcgagttcgagtcaccaaaagaagaagcggaaagtcgacaagaagtac agcatcgccgctggacatcgccaccaactctgtgggctgggcccgtgatcaccgacgagtacaaggtgcccag caagaattcaaggtgctgggcaacaccgacccggcagcagcatcaagaagaacctgatcgagccctgctgtt cgacagcggcgaaacagccgagccacccggctgaagagaaccgccagaagaagatacaccagacgga agaaccggatctgctatctgcaagagatcttcagcaacgagatggccaaggtggacgacagcttcttcacag actggaagagtccttctggtggaagaggataagaagcacgagcggcaccatcttcggcaacatcgtgga cgaggtggcctaccacgagaagtacccaccatctaccacctgagaaagaaactggtggacagcaccgaca aggccgacctgcccgtgatctatctggccctggcccacatgatcaagttccggggccacttctgatcgaggg cgacctgaaccccgacaacagcagcgtggacaagctgttcacagctggtgcagacctacaaccagctgtt cgaggaaaaccccatcaacgccagcggcgtggacgccaagggccatcctgtctgcagactgagcaagagc agacggctggaaaatctgatcgcccagctggccggcgagaagaagaatggcctgttcggaaacctgattgcc ctgagcctgggctgacccccaaactcaagagcaacttcgacctggccgaggatgccaaactgcagctgagc aaggacacctacgacgacacctggacaacctgctggcccagatcggcgaccagtagccgacctgtttctg ggcccaagaacctgtccgacgcatcctgctgagcgacatcctgagagtgaacaccgagatcaccaggc ccccctgagcgctctatgatcaagagatacagcagcaccaccaggacctgacctgtgaaagctctcgt gcggcagcagctgctgagaagtacaaagagattttctcgaccagagcaagaacggctacgcccggctacat tgacggcgaggacccagccagggaagagttctacaagttcatcaagccatcctggaaaagatggacggcaccg aggaaactgctgtaagctgaacagagaggacctgctgcggaagcagcggaccttcgacaacggcagcat ccccaccagatccacctgggagagctgcacgcatctgctggcgaggaggaagattttaccattcctgaag gacaacgggaaaagatcgagaagatcctgaccttcgcatccctactacgtggccctctggccagggg aaacagcagattcgctggtgacagaaagagcgaggaaaccatccccctggaacttcgaggaagtgg tggaagggcgcttcgccagagcttcacgagcggtgacaaactcgataagaacctgcccaacgaga aggtgctgcccaagcacagcctgctgtacgagttcaccgtgtataacgagctgacaaagtgaatactg gaccgaggggaatgagaaagccgccttctgagcggcgagcagaaaaaggccatcgtgacctgctgttca agaccaaccggaaaagtgacctgaagcagctgaagagaggactactcaagaaaatcgagtgttcgactccg tggaatctccggcgtggaagatcggtcaacgcctccctgggcacataccacgactgctgaaaattatcaag gacaaggacttctggacaatgaggaacagaggacattctggaagatatcgtgctgacctgacactgtttg aggacagagagatgatcgaggaacggctgaaaacctatgccacctgttcgacgacaaagtgatgaagcag ctgaagcggcgagatacaccgctggggcaggctgagccggaagctgatcaacggcatccgggacaag cagtccggcaagacaatcctggatttctgaagtccgacggcttcgccaacagaaactcatgcagctgatcc acgacgacagcctgaccttaagaggacatccagaaagcccaggtgtccggccagggcgatagcctgcac gagcacattgccaatctggccggcagccccgcattaagaaggcgatcctgcagacagtgaaagtggtgga cgagctcgtgaaagtgtggccggcacaagccgagaacatcgtgatcgaaatggccagagagaaccag accaccagaaggacagaagaacagccgagagaatgaagcgatcgaagaggcgatcaagagct gggcagccagatcctgaaagaacaccccggtggaacacccagctgcagaacgagaagctgtacctgtact acctgcagaatggcgggatgtacgtggaccaggaactggacatcaaccggctgtccgactacgatgtgg acgctatcgtcctcagagctttctgaaggacactccatcgacaacaaggtgctgaccagaagcgacaaga accggggcaagagcgacaacgtgcctccgaagaggtcgtgaagaagatgaagaactactggcgagct gctgaacgccaagctgattaccagagaaggttcgacaatctgaccaaggccgagagaggcgccctgagc	125

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	gaactggataaggccggcttcatcaagagacagctggtggaacccggcagatcacaaagcacgtggcaca gatcctggactcccggatgaactaagtagcagcagagaatgacaagctgatccgggaagtgaagtgtacac cctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaagtgcgcgagatcaacaactacc accacgcccacgacgctacctaagcgcgctggtgggaaccgacctgatcaaaaagtaccctaagctggaa agcgagttcgtgtacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcagcaggaat cggcaaggctaccgccaagtacttcttacagcaacatcatgaacttttcaagaccgagattacacctggccaa cggcgagatccggaagcggcctctgatcgagacaaacggcgaaaccggggagatcgtgtgggataaggg ccgggattttgccaccgtgcggaaggtctgagcatgccccaaagtgaatatcgtgaaaaagaccgaggtgca gacagcggttctagcaaaaggtctatctgccccaaaggaacagcgataagctgatcgccagaaagaagg actgggacctaagaagtacggcggtctgacagccccaccgtggcctattctgtgtggtggtggccaaagt ggaaaaggcgaagtccaagaaactgaagagtgtaaaagctgctggggatcaccatcatggaagaagc agcttcgagaagaatcccatcgacttttgaagccaagggtacaaagaagtgaagaaggacatgatcatca agctgcctaagtactccctgttcgagctggaaaacggcgggaagagaatgctggcctctgccggcgaactgc agaagggaacgaactggcctgcccctcaaatatgtgaacttctgtacgtggccagccactatgagaagct gaagggtcctccccaggataatgagcagaacagctgtttgtggaacagcacaagcactacctggacgaga tcacgagcagatcagcgagtttccaagagagtgatcctggccgacgctaacttggaacaaagtgtgtccgc ctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccacctgtttaccttgaccaat ctgggagccctgcccgttcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccaaaga ggtgtcggagccaccctgatccaccagagcatcaccggcctgtacgagacacggatcgacctgtctcagct gggaggtgacTCCGGCGGAAGCTCTGGTGGCAGCAAGCGGACCGCCGA CGGCTCTGAATTCGAGAGCCCTAAGAAGAAAAGAAAGGTGAGCG GAGGCTCTAGCGGCGGAAGCCCGAAGCGCCGGACTGCACGAAAG CAACTGCCAAAACGGACTATACAAGAAGTCCTGGAAGAACAAG CGAAGATGAGGATCGCGAAGCCAAGCGCAAGAAAAGAGGAAGAGG AAGAAGAGACTCCAAAGGAGTCCTTGACCGAAGCAGAAGTCGCA ACGGAGAAGGAAGGTGAGGATGGGGATCAGCCAACAACCCCGCC TAAACCTCTGAAAACCTCTAAGGCGGAGACACCAACTGAGAGTGT CAGCGAACCGGAGGTAGCCACGAAACAAGAGCTTCAGGAGGAAG AAGAACAGACAAAGCCACCTCGGCGGGCTCCCAAAACCTTAGCT CCTTCTTCACGCCTCGAAAGCCAGCAGTGAAGAAAGAAGTGAAGG AGGAGGAACCTGGCGCCCCTGGAAAGGAGGGCGCAGCCGAGGGC CCGCTGGACCCCTTCAGGGTATAACCCGGCAAAAAATAATTACCAC CCGGTCGAGGACGCTTGTGGAACCAAGGCCAAAAGGTACCTTAC CTCGCCGTCGCTAGGACCTTTGAGAAGATAGAGGAAGTTAGTGCT AGGTTGAGAATGGTCGAAACCTTAGTAACCTTCTCAGGTCCGTA GTGCGCCCTTAGTCCCCCAGACCTGCTTCCGGTGCTGTACCTGTCCC TGAACCATCTCGGTCCCCCACAAGGGACTGGAGTTGGGCGTCG GTGACGGCGTTCTCCTGAAAGCGGTTGCACAAGCTACAGGAAGGC AACTGGAATCTGTCCGGGCTGAGGCTGCAGAGAAAGGTGACGTGG GGCTTGTGGCAGAGAATAGTTCGGTCAACACAGCGGCTGATGCTGC CACCGCCCCCGCTTACGGCTAGTGGGGTATTCTCCAAATTTAGAG ATATAGCACGGCTGACGGGATCAGCTTCCACTGCGAAGAAGATCG ATATCATTAAGGGTTTGTTCGTGGCTTGCAAGCATTCCGAAGCACG CTTCATTGCACGCTCCCTTTCAGGGAGACTCAGACTTGGGCTGGCC GAGCAATCTGTACTGGCGGCCCTGTCTCAGGCGGTGAGCCTTACG CCGCCCGGGCAAGAGTTCCCTCCTGCGATGGTCGATGCTGGGAAG GGAAAAACCGCCGAAGCTCGAAAAACATGGCTGGAGGAGCAAGG AATGATTTTGAAGCAGACGTTCTGTGAAGTACCGGACTTGGATCG CATCATACCTGTGCTTCTCGAACATGGTTTGGAGCGGCTCCCCGAG	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	CATTGCAAACCTCTCTCCGGGCATCCCCCTCAAGCCAATGCTCGCCC ACCCCACGCGCGGAATCAGTGAGGTACTGAAACGCTTTGAAGAGG CAGCGTTTACTTGTGAATACAAGTACGATGGCCAAAGGGCACAAA TTCATGCACTTGAAGGCGGGGAAGTTAAGATATTCAGCAGGAATC AGGAGGACAACACGGGAAAATATCCTGACATAATATCTCGAATCC CTAATAATTAAGTTGCCTAGCGTAACCAGCTTCATCCTGGATACCG AAGCCGTGGCGTGGGATAGGGAGAAAAAGCAAATACAGCCATTT CAGGTGCTTACAAC TAGAAAACGAAAAGAGGTGGACGCTAGTGA AATCCAAGTCCAGGTATGTCTTTATGCCTTCGATTTGATATACCTT AATGGTGAGTCCCTTGTACGGGAACCGCTTAGTAGGAGGCGGCAG TTGCTGAGGGAAAATTTTGTGAACTGAGGGAGAGTTTGTATTT GCAACGTCATTGGATACAAAGGACATAGAACAAATAGCAGAATTT CTGGAGCAGTCAGTAAAAGACTCCTGCGAGGGCCTGATGGTGAAA ACTCTTGATGTGGACGCCACTTATGAAATCGCAAAAAGGTCACAC AATTGGCTGAACTTAAAAAGGATTACTTGGACGGGGTTCGGGGAT ACCCTCGATCTCGTCGTAATCGGAGCTTATCTCGGTAGGGGGAAG CGAGCCGGGCGATACGGAGGCTTTCTCTTGGCTAGTTATGACGAA GATTCCGAAGAGCTGCAGGCCATATGCAAGCTTGGAACGGGTTTC AGCGATGAGGAATTGGAGGAGCATCATCAGAGCTTGAAGGCACT GGTGCTCCCTCTCCTAGGCCGTACGTTAGAATAGACGGAGCAGT GATACCCGATCATTGGCTCGATCCGTCAGCTGTTTGGGAGGTGAA GTGTGCAGACCTGTCCCTCTCTCCTATTTACCCTGCAGCACGCGGT CTGGTTGACTCTGACAAAGGGATTAGCTTGAGGTTCCTAGATTTA TTCGGGTGCGCGAAGATAAACAGCCTGAACAGGCGACAACGTCCG CGCAGGTGCGCATGCCCTTATCGAAAACAGAGTCAGATCCAGAATC AACAAAGGAGAAGATTACGGGAGTGACCCGGAGGACACTTATAGT GGCGGCTCAAAACGAACCGCCGATAGTCAGCATTCAACACCTCCA AAAACTAAAAGGAAAGTCGAGTTTGAGCCAAAGAAGAAGCGCAA AGTCTAA	
NLS- nCas9- linker- hLIG1(1 19-919)- bpNLS	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDE VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYFLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV DELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIIEGKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHV	126

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIN NYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIK SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGI TIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRKML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHHRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSTGLYETR IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSPKRR TARKQLPKRTIQEVLEEQSEDEDEAKRKKEEEEEETPKESL TEAEVAT EKEGEDGDQPTTPPKPLKTSKAETPTESVSEPEVATKQELQEEEEQTK PPRRAPKTLSSFTPRKPAVKKEVKEEPPGAPGKEGAAEGPLDPSGYN PAKNNYHPVEDACWKPGQKVPYLA VARTFEKIEEV SARLRMVETLS NLLRSVVALSPPDLLPVLYLSLNLHGPPQQGLELGVDGVLLKAVAQ ATGRQLESVRAEAAEKGDVGLVAENSRSTQRLMLPPPPLTASGVFSK FRDIARLTGSASTAKKIDIIKGLFVACRHSEARFIARSLSGRLRLGLAE QSVLAALSQAVSLTPPGQEFPPAMVDAGKGKTA EARKTWLEEQGM ILKQTFCEVPDLDRIPVLLHGLERLPEHCKLSPGIPLKPM LAHPTRGIS EVLKRFEAAFTCEYKYDQGQRAQIHAELEGGEVKIFSRNQEDNTGKYP DIISRIPIKLP SVTSFILDTEAVAWDREKKQIQPFQVLTTRKRKEVDAS EIQQVQCLYAFDLIYLN GESLVREPLSRRRQLLRENFVETEGEFVFATS LDTKDIEQIAEFLEQSVKDSCEGLMVKTLDVDATYEIAKRSHNWLKL KKDYLDGVGDTLDLVVIGAYLGRGKRAGRYGGFLLASYDEDESEELQ AICKLGTGFSDEELEHHQSLKALVLPSPRPYVRIDGAVIPDHWLDP AVWEVKCADLSLSPIYPAARGLVDSKGISLRFPRFIRVREDKQPEQA TTSAQVACLRYRKQSQIQNQGGEDSGSDPEDTYSGGSKRTADSQHSTP PKTKRKVEFEPKKKKRKV*	
NLS- nCas9- linker- hLIG1(2 33-919)- bpNLS	atgaaacggacagccgacggaagcgagtcgagtcaccaaagaagaagcggaaagtcgacaagaagtac agcatcggcctggacatcggcaccactctgtgggctgggcccgtgatcaccgacgagtagaaggtgcccag caagaattcaaggtgctgggcaacaccgacggcagcagcatcaagaagaacctgatcgagccctgctgtt cgacagcggcgaaacagccgagggccaccggctgaagagaaccgccagaagaagatacaccagacgga agaaccggatctgctatctgcaagagatcttcagcaacgagatggccaaggtggacgacgcttctccacag actggaagagtcctcctggtggaagaggataagaagcagcagcggcaccatcttcggcaacatcgtgga cgaggtggcctaccacgagaagtagccaccatctaccacctgagaagaactggtggacagcaccgaca agggcgacctgcccgtgatctatctggccctggcccatgatcaagttccggggccacttctgatcagggg cgacctgaaccccgacaacagcagcgtggacaagctgttcacagctggtgcagacctacaaccagctgtt cgaggaaaaccccatcaacgccagcggcgtggacgccaagccatcctgtctgccagactgagcaagagc agacggctggaaaatctgatcggcagctggcggcgagaagaagaatggcctgttcggaacctgattgcc ctgagcctgggctgaccccaacttcaagagcaacttcgacctggcggaggtgccaactgcagctgagc aaggacacctacgacgacacctggacaacctgctggccagatcgccgaccagtagccgacctgttctg gccgcaagaacctgtccgacgcatcctgtgagcgacatcctgagagtgaacaccgagatcacaaggc cccctgagcgctctatgatcaagagatagcagcaccaccaggacctgacctgtgaaagctctcgt gcggcagcagctgcctgagaagtacaaagagattttctcgaccagagcaagaacggctacgcccgtacat tgacggcggagccagccaggaaggttctacaagttcatcaagccatcctggaaaagatggacggcaccg aggaactgctcgtgaagctgaacagagaggacctgctcggaagcagcggaccttcgacaacggcagcat cccaccagatccacctgggagagctgcacgccattctgcggcggcaggaagattttaccattcctgaag gacaacggggaagatcgagaagatcctgacctccgcatccctactacgtgggcccctctggccagggg aaacagcagattcgctggatgaccagaaagagcaggaagaaaccatcaccctggaacttcgaggaagtgg	127

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	tggacaagggcgcttccgccagagcttcacgagcggtatgaccaacttcgataagaacctgcccacgaga aggtgctgcccagcacagcctgtgtacgagtacttcaccgtgtataacgagctgaccaaagtgaatacgt gaccgaggggaatgagaagccccgcttctgagcggcgagcagaaaaaggccatcgtggacctgctgttca agaccaaccggaaagtgaccgtgaagcagctgaagaggactacttcaaaaaatcgagtgttcgactccg tggaaatctccggcggtggaagatcggttcaacgctccctgggcacataccacgatcgtgtaaaattatcaag gacaaggacttctggacaatgaggaacgaggacattctggaagatatcgtgctgacctgacactgtttg aggacagagagatgatcaggaacggctgaaaacctatgccacctgttcgacgacaaagtgatgaagcag ctgaagcggcgagataccggctggggcaggctgagccggaagctgatcaacggcatccgggacaag cagtcgggaagacaatcctggatttctgaagtccgacggcttcgccaacagaaacttcatgcagctgatcc acgacgacagcctgacctttaaaggacatccagaaagcccaggtgtccggccaggcgcatagcctgcac gagcacttgccaatctggccggcagccccgccattaagaaggccatcctgcagacagtgaaggtggtgga cgagctcgtgaaagtgatggccggcacaagcccagaaacatcgtgatcgaatggccagagagaaccag accaccagaaagggacagaagaacagccgagagagaatgaagcggatcgaagaggccatcaagagct gggcagccagatcctgaaagaacccccgtggaaaaccccagctgcagaacgagaagctgtacctgtact acctgcagaatggcggggatgtacgtggaccaggaactggacatcaaccggctgtccgactacgatgtgg acgctatcgtcctcagagcttctgaaggacgactccatcgacaacaaggtgctgaccagaagcgacaaga accggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaactactggcgcgagct gctgaacgccaagctgattaccagagaaagttcgacaatctgaccaaggccgagagaggcgccctgagc gaactggataagggcggttcatcaagagacagctggtggaaacccggcagatcacaaagcacgtggcaca gatcctggactcccgatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaagtgatcac cctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaagtgcgcgagatcaaaactacc accacgcccacgacgctacctaagcgcgtggtggaaccgccctgatcaaaaagtaccctaagctggaa agcgagttcgttacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcgagcaggaat cggcaaggctaccgccaagtacttctctacagcaacatcatgaacttttcaagaccgagattacctggccaa cggcgagatccggaagcggcctctgatcgagacaaacggcgaaacccggggagatcgtgtgggataaggg ccgggattttgccaccgtgcggaaagtgtgagcatccccaagtgaatatcgtgaaaaagaccgaggtgca gacagggcggtttagcaaaagagctatctgcccagaggaacagcgataagctgatcgccagaaagaagg actgggaccttaagaagtacggcggttgcagacgcccaccgtggcctattctgtgctggtggtggccaaagt ggaaaagggcaagtccaagaactgaagagtgtaagagctgctggggatcacatcatggaagaagc agcttcgagaagaatccatcgactttctggaagccaagggtacaaagaagtgaagaaggacatgatcatca agctgcctaagtactccctgttcgagctggaaaacggccggaagagaatgtggcctctgcccggcgaactgc agaagggaacgaactggcctgcccctccaatatgtgaacttctgtacgtggccagccactatgagaagct gaagggtcccccgaggataatgagcagaacagctgtttgtggaacagcacaagcactacctggacgaga tcatcgagcagatcagcaggttctcaagagagtgatcctggccgacgctaacttggaacaaagtgtgtccgc ctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccacctgtttacctgaccaat ctgggagcccctgcccctcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccaaaga ggtgtgagcgccaccctgatccaccagagcatcaccggcctgtacgagacaggatcgacctgtctcagct gggaggtgacTCCGGCGGAAGCTCTGGTGGCAGCAAGCGGACCGCCGA CGGCTCTGAATTTCGAGAGCCCTAAGAAGAAAAGAAAGGTGAGCG GAGGCTCTAGCGGCGGAAGCACACCCAGGAAACCAGCCGTGAAA AAAGAGGTTAAAGAAGAGGAACCTGGGGCTCCGGGAAGGAGGG AGCAGCGGAAGGTCCGCTCGACCCTTCAGGATACAACCCAGCCAA AAACAACCTACCACCCCGTAGAGGATGCTTGCTGGAAGCCAGGCCA AAAGGTGCCCTATTTGGCCGTTGCTAGGACTTTCGAAAAAATTGA GGAGGTGAGCGCGCGACTCAGAATGGTAGAGACTCTGTCTAACCT CCTTCGCTCCGTAGTGGCTCTTTCACCTCCAGATCTTCTTCCAGTGC TGTACCTGAGCCTGAACCACTTGGGCCCTCCCCAGCAGGGACTGG AACTGGGCGTAGGGGACGGAGTATTGCTGAAGGCTGTTGCTCAGG CAACCGGACGACAGCTCGAGTCTGTGCGAGCAGAAGCTGCAGAA	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	AAGGGGGACGTCGGGTTGGTTGCCGAAAATTCAAGATCTACCCAA CGATTGATGTTGCCACCGCCGCCTCTGACTGCGTCAGGTGTATTCT CCAAGTTCGGGATATTGCCAGGCTTACGGGTAGCGCTTCCACTG CTAAAAAGATCGACATAATAAAAGGTCTGTTTCGTCGCTTGTCCGCC ATTCAGAGGCGAGGTTTATAGCCAGATCCCTTTCCGGACGACTTC GACTCGGCTTGGCTGAGCAGTCAGTACTGGCAGCTTTGTCTCAAG CTGTATCACTCACGCCCCCGGACAAGAATTTCCACCCGCCATGGT TGACGCAGGCAAGGGTAAGACTGCTGAGGCAAGAAAGACGTGGC TGGAGGAACAAGGTATGATACTTAAACAAACGTTTTGCGAAGTTC CGGACTTGGACCGGATCATACCTGTGTTGCTGGAGCACGGCCTCG AGCGCTTGCCCGAACACTGTAACTGTCTCCAGGAATACCTCTCA AACCCATGTTGGCTCATCCTACGAGGGGAATCTCAGAGGTACTTA AACGGTTTGAAGAAGCCGCTTTCACGTGCGAATACAAGTATGATG GTCAGAGAGCGCAAATCCACGCATTGGAAGGGGGTGAGGTAAAG ATTTTTTCAAGGAATCAGGAGGACAATACAGGGAAGTACCCCGAT ATCATCAGTCGGATTCTTAAATTAAGCTTCCATCAGTCACGTCTCT TCATACTGGACACTGAGGCAGTGGCTTGGGACCGAGAGAAGAAG CAGATACAACCCTTTCAGGTACTTACAACCAGAAAGCGCAAGGAA GTCGACGCTTCTGAGATTCAAGTACAAGTCTGCCTTTATGCGTTTG ACCTGATCTATCTTAATGGAGAGAGTTTGGTGAGAGAACCCTTGA GCAGACGACGGCAGCTCTTGAGAGAAAATTTCTAGAAAAGTGAAG GGGAGTTTCGTCTTTGCGACTAGTCTCGACACCAAAGACATTGAGC AAATCGCGGAATTCCTCGAACAGTCAGTTAAAGACTCCTGCGAAG GTCTGATGGTTAAGACTCTTGACGTGGATGCTACCTACGAGATAG CTAAGCGGTCACACAATTGGCTGAAACTGAAAAAGGACTATCTGG ATGGAGTTGGGGACACGCTGGATTGTTGTCGTTATCGGGGCCTATC TGGGACGCGGTAAGCGGGCAGGGAGATATGGTGGATTCTCTCCTCG CTTCATACGATGAGGACTCTGAAGAGCTGCAGGCTATATGCAAAC TTGGGACGGGTTTTTCCGATGAAGAATTGGAGGAACATCATCAGT CACTGAAGGCCCTTGTATTGCCAAGTCCACGCCCATACGTACGAA TCGATGGAGCAGTAATCCCTGACCACTGGCTTGACCCGTCGCCCG TCTGGGAAGTAAAGTGCAGCGGATCTCTCTCAGTCCGATCTACCC AGCCGCACGGGGGCTGGTTGACAGTGACAAGGGTATCAGCCTGCG ATTTCTCGATTCTACGCGTCCGGGAAGACAAGCAACCGGAACA GGCTACGACCTCTGCACAGGTCGCATGTTTGTATAGAAAACAGAG CCAAATTCAGAATCAACAAGGCGAAGACAGTGGGTCCGATCCTGA AGATACCTACTCAGGCGGCAGTAAACGGACAGCTGATAGCCAACA CTCAACTCCTCCGAAGACTAAAAGGAAGGTAGAGTTCGAACCAAA AAAGAAAAGGAAAGTGTA	
NLS- nCas9- linker- hLIG1(2 33-919)- bpNLS	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAKVDSSFFHRLEESFLVEEDKKHERHPIFGNIVDE VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYFLKDNREKIEKILTFRIPYYVGPLARGNS	128

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQML IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIIEGKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIN NYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVDVRKMIK SEQIIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGI TIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRKML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQ HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSTGLYETR IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSTPRKP AVKKEVKEEPEGAPGKEGAAEGPLDPSGYNPAKNNYHPVEDACWKP GQKVYPYLA VARTFEKIEEVSARLRMVETLSNLLRSVVALSPPDLLPVL YLSLNLHGPPQQGLELGVGDGVLLKAVAQATGRQLESVRAEAAEK DVGLVAENSRSTQRLMLPPPPLTASGVFSKFRDIARLTGSASTAKKIDI IKGLFVACRHSEARFIARSLSGRLRLGLAEQSVLAALSQAVSLTPPGQ EFPAMVDAGKGKTAEARKTWLEEQGMILKQTFCEVPDLDRIPVLL EHGLERLPEHCKLSPGIPLKPMLAHPTRGISEVLKRFEEAAFTCEYKY DGQRAQIHAELEGGEVKIFSRNQEDNTGKYPDIISRIPKIKLPSVTSFILD TEAVAWDREKKQIQPFQVLTTRKRKEVDASEIQVQVCLYAFDLIYLN GESLVREPLSRRRQLLRENFEVTEGEFVFATSLDTKDIEQIAEFLEQSV KDSCEGLMVKTLDVDATYEIAKRSHNWLKLLKDYLDGVDGDTLDLV VIGAYLGRGKRAGRYGGFLLASYDEDSEELQAICKLGTGFSDEELEE HHQSLKALVLPSPRPYVRIDGAVIPDHWLDPSAVWEVKCADLSLPIY PAARGLVDSKGISLRFPRFIRVREDKQPEQATTSAQVACLRYKQSQI QNQQGEDSGSDPEDTYSGGSKRTADSQHSTPPKTKRKVEFEPKKRK V*	
NLS- nCas9- linker- SplintR- bpNLS	atgaaacggacagccgacggaagcgagttcgagtcaccaaagaagaagcggaaagtcgacaagaagtac agcatcgccctggacatcgccaccaactctgtgggctgggcccgtgatcaccgacgagtagaaggtgcccag caagaattcaaggtgctgggcaacaccgaccggcacagcatcaagaagaacctgatcgagccctgctgtt cgacagcggcgaaacagccgagggccaccggctgaagagaaccgccaagaagaagatacaccagacgga agaaccggatctgctatctgcaagagatctcagcaacgagatggccaaggtggacgacagcttctccacag actggaagagtccttctggtggaagaggataagaagcagcagcggcaccatcttcggcaacatcgtgga cgaggtggcctaccacgagaagtacccaccatctaccacgtgagaaagaaactggtggacagcaccgaca aggccgacctgcggtgatctatctggccctggcccatgatcaagttccggggccacttctgatcagggg cgacctgaacccgacaacagcagctggacaagctgttcacagctggtgcagacctacaaccagctgtt cgaggaaaacccatcaacgccagcggcgtggacccaagccatcctgtctgccagactgagcaagagc agacggctggaataatctgatcggcagctgcccggcgagaagaagaatggcctgttcggaacctgattgcc ctgagcctgggctgaccccaactcaagagcaacttcgacctggcgaggatgccaactgcagctgagc aaggacacctacgacgacgacctggacaacctgctggccagatcgccgaccagtagcggacctgtttctg	129

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	gcccgaagaacctgtccgacgccatcctgctgagcgacatcctgagagtgaacaccgagatcaccaaggc ccccctgagcgctctatgatcaagagatacgagcaccaccaggacctgacctgctgaaagctctcgt gcggcagcagctgcctgagaagtacaaagagattttctcgaccagagcaagaacggctacgcccggctacat tgacggcggagccagccaggaagagtctacaagttcatcaagccatcctggaaaagatggacggcaccg aggaaactgctcgtgaagctgaacagagaggacctgctcggaagcagcggaccttcgacaacggcagcat ccccaccagatccacctgggagagctgcacgccattctgcggcggcaggaagattttaccattcctgaag gacaaccgggaaaagatcgagaagatcctgaccttcgcacccctactacgtgggcccctctggccagggg aaacagcagattcgctggatgaccagaaaagcagcaggaaaccatcacccctggaaactcgaggaaagtg tggacaaggcgcttcgccagagcttcacgagcggatgaccaacttcgataagaacctgcccacagaga aggtgctgcccaagcacacctgctgtacgagtacttcaccgtgtataacgagctgacaaagtgaatacgt gaccgagggaaatgagaagccccgcttctgagcggcagcagaaaaaggccatcgtggacctgctgttca agaccaaccggaaagtgacctgaagcagctgaaagaggactacttcaaaaaatcgagtgttcgactccg tggaaatctccggcgtggaagatcggttaacgcctccctgggcacataccacgatcgtgtaaaattatcaag gacaaggacttctggacaatgaggaacaggacattctggaagatcgtgctgacctgacactgtttg aggacagagagatgatcgaggaacggctgaaaacctatgccacctgttcgacgacaaaagtgatgaagcag ctgaagcggcggagataccggcgtggggcaggctgagccggaagctgatcaacggcatccgggacaag cagtcgggcaagacaatcctgatttctgaagtccgacggcttcgccaacagaaaactcatgcagctgatcc acgacgacagcctgacctttaagaggacatccagaaagcccagggttcggccaggggcgatagcctgcac gagcacttggcaatctggcggcagccccgccattaagaaggcgatcctgcagacagtgaaggtggtgga cgagctcgtgaaagtgtggcggcagcaagcccagagaacatcgtgatgaaatggccagagagaaccag accaccagaaagggacagaagaacagccgagagagaatgaagcggatcgaagaggcgatcaagagct gggcagccagatcctgaaagaacccccgtggaacacccagctgcagaacgagaagctgtacctgtact acctgcagaatggcggggatgtacgtggaccaggaactggacatcaaccggctgtccgactacgatgtgg acgtatcgtgctcagagcttctgaaggacgactccatcgacaacaaggtgctgaccagaagcgacaaga accggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaactactggcggcagct gctgaacgccaagctgattacccagagaaagttcgacaatctgaccaaggccgagagaggcgccctgagc gaactggataaggccggttcatcaagagacagctggtggaaacccggcagatcacaaagcagctggcaca gatcctggactcccggatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaagtgatcac cctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaagtgcgcgagatcaacaactacc accacgcccacgacgcctacctaagcgcctgctgggaaccgccctgatcaaaaagtaccctaagctggaa agcaggttcgtgtacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcagcaggaat cggcaaggctaccgccaagtacttcttacagcaacatcatgaacttttcaagaccgagattacctggccaa cggcgagatccggaagcggcctctgatcgagacaaacggcgaacccggggagatcgtgtgggataagg ccgggattttgccaccgtgcggaagtgtgagcatgccccaaagtgaatatcgtgaaaaagaccgaggtgca gacaggcggcttcagaaagagtctatcctgccaagaggaacagcgataagctgatcgccagaaagaagg actgggaccctaagaagtacggcggcttcgacagccccaccgtggcctattctgtgctggtggtggccaaagt ggaaaagggaagtccaagaactgaagagtgtaagagctgctggggatcaccatcatggaagaagc agcttcgagaagaatccatcgacttctggaagccaagggtacaaagaagtgaagaaggacctgatcatca agctgcctaagtactccctgttcgagctggaaaacggccggaagagaatgctggcctctgcggcgaaactgc agaagggaacgaactggcctgcctccaatatgtgaacttctgtacctggccagccactatgagaagct gaagggtcccccgaggataatgagcagaacagctgtttgtggaacagcacaagcactacctggacgaga tcacgagcagatcagcaggttctccaagagagtgatcctggccgacgctaactggacaaagtgtgtccgc ctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccactgtttaccctgaccaat ctgggagcccctgcgccttcaagtactttgacaccaccatcgaccggaagaggtacaccagaccaaaga ggtgtcggacgccaccctgatccaccagagcatcaccggcctgtacgagacagggatcgacctgtctcagct gggaggtgacTCCGGCGGAAGCTCTGGTGGCAGCAAGCGGACCGCCGA CGGCTCTGAATTTCGAGAGCCCTAAGAAGAAAAGAAAGGTGAGCG GAGGCTCTAGCGGCGGAAGCATGGCAATCACTAAGCCCCCTCTTGG CGGCGACTTTGGAAAACATCGAGGATGTGCAATTCCCGTGCCTTG	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	CCACACCAAAGATAGACGGGATCCGATCAGTGAAGCAAACGCAG ATGCTCTCTAGAACGTTCAAGCCTATTAGAACTCAGTGATGAAT CGGCTCTTGACTGAGCTGTTGCCGGAAGGCAGCGATGGGGAAATA TCTATCGAGGGAGCCACATTTCAGACACTACGAGCGCCGTAATG ACTGGACATAAGATGTATAATGCTAAATTCTCCTACTATTGGTTTG ACTATGTCAGTATGACCTCTTAAGAAATATATAGACCGAGTGG AGGATATGAAAAATTATATTACTGTACACCCGCATATTCTGGAAC ATGCCCAAGTTAAGATTATTCCTCTCATTCCCGTCGAGATTAATAA TATCACAGAACTGCTTCAGTATGAGCGCGACGTATTGTCCAAAGG CTTTGAAGGGGTTATGATACGCAAACCGGACGGCAAGTACAAGTT CGGAAGAAGCACATTGAAAGAGGGGTATATTGCTGAAGATGAAGC AGTTTAAGGATGCTGAGGCAACAATAATCAGCATGACAGCACTTT TAAAAAATACCAACACGAAAATAAGGACAATTTTGGTTATAGTA AGCGGTCAACGCACAAAAGTGGGAAGGTAGAAGAAGACGTAATG GGTAGCATTGAGGTGGATTATGACGGGGTGGTTTTTCAGCATAGGG ACTGGGTTTGATGCAGATCAACGGAGGGACTTTTGGCAGAACAAA GAATCATATATAGGCAAAATGGTAAAGTTCAAATACTTCGAAATG GGAAGTAAAGACTGCCCCAGATTCCCTGTATTTCATTGGCATCAGG CACGAGGAGGACAGGAGTGGGGGATCAAAGCGGACTGCTGATAG TCAGCATAGTACTCCACCCAAGACCAAGCGGAAAGTTGAGTTTGA GCCGAAGAAAAAGCGAAAAGTGTA	
NLS- nCas9- linker- SplintR- bpNLS	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAK VDDSFHRLSEESFLVEEDKKHERHPIFGNIVDE VAYHEKYPTIYHLRKKLV DSTDKADLR LIYLA LAHMIKFRGHFLIEG DLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNLGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEFVVDKGASQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFML IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAICKGILQTVKVV DELVKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIIEGKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRIN NYHHAHDAYLNAVVGTA LIKKY PKLESEFVYGDYKVYDVRKMIK SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGI TIMERSSEKPNIDFLEAKGYKEVKKDLIKLPKYSLELENKRKML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENII	130

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	HLFTLTNLGAPAAFKYFDDTTIDRKRYTSTKEVLDTLHQSIITGLYETR IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGSSGGSMATK PLLAATLENIEDVQFPCLATPKIDGIRSVKQTQMLSRTFKPIRNSVMNR LLTELLPEGSDGEISIEGATFQDTTSAVMTGHKMYNAKFSYYWFDYV TDDPLKKYIDRVEDMKNYITVHPHILEHAQVKIHLIPVEINNITELLQY ERDVL SKGFEGVMIRKPDGKYKFGRSTLKEGILLKMKQFKDAEATIIS MTALFKNTNTKTKDNFGYSKRSTHKS GKVEEDVMGSIEVDYDGVVF SIGTGFDADQRRDFWQNKESYIGKMVKFKYFEMGSKDCPRFPVFIGI RHEEDRSGGSKRTADSQHSTPPKTKRKVEFEPKKRKRV*	
NLS- nCas9- linker- T4LIG- bpNLS	atgaaacggacagccgacggaagcaggttcagtcaccaagaagaagcggaaagtcgacaagaagtac agcatcggcctggacatcggcaccactctgtgggctggccgctgatcaccgacgagtacaaggtgccag caagaaattcaaggtgctgggcaacaccgaccggcacagcatcaagaagaacctgatcggagccctgctgtt cgacagcggcgaaacagccgagggccaccggctgaagagaaccgcgaagaagatacaccagacgga agaaccggatctgctatctgcaagagatcttcagcaacgagatggccaaggtggagcagacagcttctccacag actggaagagtccttctgtggaagaggataagaagcagcagcggcaccatcttcggcaacatcgtgga cgaggtggcctaccacgagaagtacccaccatctaccacctgagaagaagaactgggtggacagcaccgaca aggcggacctgcggtgatctatctggccctggccacatgatcaagttccggggccacttctgatcgaggg cgacctgaacccgacaacagcgacgtggacaagctgttcacagctgtgtgcagacctacaaccagctgtt cgaggaaaacccatcaacggcagcggcggtggacgccaaggccatcctgtctgccagactgagcaagagc agacggctggaaaatctgatcgccagctgcccggcgagaagaagaatggcctgttcggaacacctgattgcc ctgagcctgggctgaccccaacttcaagagcaacttcgacctggccgaggtatgccaactgcagctgagc aaggacacctacgacgacacctggacaacctgtggccagatcggcgaccagtacgccacctgtttctg gccgccaagaacctgtccgacgccatcctgtctgagcgacatcctgagagtgaacaccgagatcaccaaggc ccccctgagcgcctctatgatcaagagatacagcagcaccaccaggacctgacctgtctgaagctctcgt gcggcagcagctgcctgagaagtacaagagattttctcgaccagagcaagaacggctacgcccggctacat tgacggcgaggccagccaggaagagtctacaagttcatcaagccatcctggaaaagatggacggcaccg aggaaactgctcgtgaagctgaacagagaggacctgtcgggaagcagcggaccttcgacaacggcagcat ccccaccagatccacctgggagagctgcacgccattctgcggcggcaggaagattttaccattcctgaag gacaacccgggaaaagatcgagaagatcctgacctccgcatccccctactacgtgggcccctctggccagggg aaacagcagattcgctggatgaccagaagagcgaggaaacatcacccccctggaacttcgagggaagtgg tggacaaggcgcttcgcccagagcttcacgagcggatgaccaacttcgataagaacctgcccacagaga agggtgctgcccagcagcctgtgtacgagtaattcaccgtgtataacgagctgaccaaagtgaatactgt gaccgagggaatgagaaaagcccgccttctgagcggcgagcagaaaaaggccatcgtggacctgctgttca agaccaaccggaaaagtaccgtgaagcagctgaaagaggactacttcaagaaaatcgagtgttcgactccg tggaaatcctccgctgggaagatcggttaacgcctccctgggacataccacgatctgctgaaaattatcaag gacaaggacttctggacaatgaggaaaaacgaggacattctggaagatatcgtgtgacctgacactgtttg aggacagagagatgatcagggaacggctgaaaacctatgcccacctgttcgacgacaaaagtgatgaagcag ctgaagcggcgagatacacccgctggggcaggctgagccggaagctgatcaacggcatccgggacaag cagtcgggcaagacaatcctgatttctgaagtcgcacggcttcgccaacagaaactcatgcagctgatcc acgacgacagcctgacctttaagaggacatccagaagcccaggtgtccggccagggcgatagcctgcac gagcacattgccaatctggccggcagccccgccaatgaagggcagctcgcagacagtgaaggtgtgtgga cgagctcgtgaaagtgtgggcccggcacaagcccgaagaacatcgtgatcgaatggccagagagaaccag accaccagaaggagacagaagaacagccgcgagagaatgaagcggatcgaagagggcacatcaagagct gggacgacagatcctgaaagaacacccgctggaaaaacacacagctgcagaacgagaagctgtacctgtact acctgcagaatgggcccggatgtgtacgtggaccagggaactggacatcaaccggctgttcgactacgatgtgg acgctatcgtgctcagagctttctgaagcagactccatcgacaacaagaggtgtgacacagaagcgacaaga accggggcaagcagcgaacctgcctccgagaagagctgtgaagaagatgaagaactactgtggcgagct gctgaacgccaagctgattaccagagaagaagtgcacaatcgtgaccaaggccgagagaggcgccgctgagc gaactggataaggccgcttcacgaagagacagctgtgggaagcccgagatcacaaagcagctggcaca	131

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	gatcctggactcccggatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaaagtgtcac cctgaagtccaagctgggtgccgatttccggaaggatttccagttttacaagtgcgcgagatcaacaactacc accacgcccacgacgcctacctgaacgccgtcgtgggaaccgccctgatcaaaaagtaccctaagtggaa agcgagttcgtgtacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcgagcaggaat cggcaaggctaccgccaagtacttcttacagcaacatcatgaacttttcaagaccgagattaccctggccaa cggcgagatccggaagcggcctctgatcgagacaaacggcgaaaccggggagatcgtgtgggataaggg ccgggattttgccaccgtcgggaaagtgtgagcatgcccaagtgaatatcgtgaaaagaccgaggtgca gacagcggttcagcaaagagtctatctcgccaagaggaaacagcgataagctgatcgccagaaagaagg actgggacctaagaagtacggcggttcgacagccccaccgtggcctattctgtgtggtggtggccaaagt ggaaaaggcgaagtccaagaaactgaagagtgtgaaagagctgctggggatcaccatcatggaagaagc agcttcgagaagaatccatcgacttttgaagccaagggtacaaagaagtgaaaaaggacctgatcatca agctgcctaagtactccctgttcgagctggaaaacggcgggaagagaatgctggcctctgccggcgaactgc agaagggaacgaactggcctgcctccaaatatgtgaacttctgtacctggccagccactatgagaagct gaagggtcccccgaggataatgagcagaacagctgtttgtggaacagcacaagcactacctggacgaga tcacgagcagatcagcgagtttccaagagagtgtcctggccgacgctaactggacaaagtgtgtccgc ctacaacaagcaccgggataagcccacagagagcaggccgagaatatcatccacctgtttaccctgaccaat ctgggagccctgccgcctcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccaaaga ggtgctggacgccacctgatccaccagagcatcaccggcctgtacgagacacggatcgacctgtctcagct gggaggtgacTCCGGCGGAAGCTCTGGTGGCAGCAAGCGGACCGCCGA CGGCTCTGAATTCGAGAGCCCTAAGAAGAAAAGAAAGGTGAGCG GAGGCTCTAGCGGCGGAAGCATGATTCTTAAATTTCTTAACGAGA TTGCGAGTATTGGCAGCACGAAACAAAAGCAGGCCATATTGGAAG AGAATAAGGACAATGAGTTGCTTAAACGCGTGTATAGGCTCACTT ACTCTCGCGGACTGCAATACTATATTAATAAATGGCCTAAGCCCG GCATCGCTACTCAAAGCTTCGGAATGCTTACGCTGACAGATATGC TCGACTTCATCGAGTTTACTCTCGCAACAAGGAAGTTGACTGGCA ACGCCGCGATTGAAGAATTGACGGGTTATATCACGGACGGGAAGA AGGATGATGTTGAGGTGCTGAGGCGCGTTATGATGCGCGACCTCG AATGTGGTGCCTCAGTTTCCATAGCCAATAAAGTTTGGCCAGGCTT GATCCCGGAGCAGCCACAGATGCTGGCCAGTAGCTACGACGAGA AGGGTATTAACAAAATATCAAGTTTCCAGCGTTTGCACAACCTTA AAGCGGATGGGGCGCGGTGTTTCGCCGAAGTCCGGGGTGACGAAT TGGACGATGTGCGCCTTCTGAGTCGCGCAGGAAATGAATATCTGG GGCTTGACCTCTTGAAGGAGGAGCTGATTAAAGATGACAGCAGAAG CCAGGCAGATCCATCCAGAGGGGGTACTTATTGATGGTGAACCTCG TATACCATGAGCAGGTTAAGAAGGAGCCAGAGGGTTTGGATTTC TCTTTGACGCCTATCCCGAGAATTCAAAGGCAAAGGAGTTCGCCG AGGTTGCAGAATCAAGAACGGCTTCCAACGGCATAGCGAATAAAT CACTCAAAGGAATATATCTGAAAAGGAGGCACAGTGTATGAAT TCCAAGTGTGGGACTATGTGCCGCTTGTGCGAGATTTACAGCTTGCC TGCTTTCCGATTGAAGTACGATGTACGGTTTAGTAAGCTCGAGCA AATGACTTCAGGTTACGATAAAGTCATCTTGATTGAGAACCAGGT CGTTAATAATCTTGACGAGGCGAAGGTCATATATAAGAAATATAT AGATCAAGGGCTCGAGGGTATCATTCTGAAGAATATAGATGGCTT GTGGGAAAACGCCAGGTCCAAAACCTGTATAAGTTTAAGGAAGT AATAGATGTAGATTTGAAAATAGTTGGAATTTACCCCCATCGGAA GGACCCACGAAAGCGGGTGGGTTTATCCTCGAGAGCGAGTGTGG GAAGATAAAAGTGAATGCCGGCTCCGGATTGAAGGACAAGGCAG GTGTGAAAAGTCATGAGCTCGATCGGACGAGAATAATGGAAAAC	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	CAGAATTACTACATTGGAAAAGATTTTGGGAATGCGAGTGTAACGGC TGGTTGAAGAGCGACGGACGCACCGATTACGTGAAACTCTTTCTG CCAATTGCAATCAGGTTGAGAGAGGATAAGACTAAGGCCAATACT TTCGAGGACGTCTTCGGGAGACTTTCACGAAGTCACTGGGCTTTCTG GGGGTAGTAAGAGAACTGCAGATAGCCAGCATTCAACGCCGCCA AAAACAAAGCGAAAGGTAGAAATTCGAACCAAAGAAAAAGCGGAA AGTATAA	
NLS- nCas9- linker- T4LIG- bpNLS	MKRTADGSEFESPKKKRKVDKKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAKVDSSFFHRLEESFLVEEDKKHERHPIFGNIVDE VAYHEKYPTIYHLRKKLV DSTDKADLR LIY LALAHMIKFRGHFLIEG DLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNLGLFNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYFLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEVVDK GASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFML IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIIEGKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYW RQLLNAKLITQRKFDNLTKAERGGSEL DKA GFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREIN NYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIK SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGI TIMERSSEFKNPIDFLEAKGYKEVKKDLIKLPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETR IDLSQLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSMILKI LNEIASIGSTKQKQAILEKNKDNELLKRVYRLTYSRGLQYYIKKWPKP GIATQSF GMLTLDMLDFIEFTLATRKL TGNA AIEELTG YITD GKKDD VEVLRRVMMRDLECGASVSIANKVWPGLIPEQPQMLASSYDEKGIN KNIKFP AFAQLKADGARCFAEVRGDELDDVRLLSRAGNEYLGDL KEELIKMTAEARQIHPEGVLIDGELVYHEQVKKEPEGLDFLFDAYPEN SKAKEFAEVAESRTASNGIANKSLKGTISEKEAQCMKFQVWDYVPLV EIYSLPAFRLKYDVRFSKLEQMTSGYDKVILIENQV VNNLDEAKVIYK KYIDQGLEGIILKNIDGLWENARSKNLYKFKEVIDVDLKIVGIYPHRK DPTKAGGFILESECGIKVNAGSGLKDKAGVKSHELDRTRIMENQNY YIGKILECECNGWLKSDGR TDYVKLFLPIAIRLREDKTKANTFEDVFG DFHEVTGLSGGSKRTADSQHSTPPKTKRKVEFEPKKKKRKV*	132

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
NLS-nCas9-LZ	atgaaacggacagccgacggaagcgagttcgagtcaccaaaagaagaagcggaaagtcgacaagaagtac agcatcgccctggacatcggcaccactctgtgggtgggcccgtgatcccgacgagtagacaaggtgccag caagaattcaaggtgctgggcaacaccgaccggcacagcatcaagaagaacctgatcggagccctgctgtt cgacagcggcgaaacagccgagggcccccggctgaagagaaccgccagaagaagatacaccagacgga agaaccggatctgctatctgcaagagatcttcagcaacgagatggccaaggtggacgacagcttctccacag actggaagagtcttctggtggaagaggataagaagcacgagcggcaccatcttcggcaacatcgtgga cgaggtggctaccacgagaagtacccaccatctaccacctgagaaagaactggtggacagcaccgaca aggccgacctgcccgtgatctatctggccctggcccacatgatcaagttccggggccacttctgatcagggg cgacctgaaccccgacaacagcagcgtggacaagctgttcatccagctggtgcagacctacaaccagctgtt cgagggaaaaccccatcaacgcccagcggcgtggacgccaagccatcctgtctccagactgagcaagagc agacggctggaaaatctgatcggccagctggcggcgagaagaagaatggcctgttcggaaaactgattgcc ctgagcctgggctgacccccaaactcaagagcaacttcgacctggccgaggatgccaactgcagctgagc aaggacacctacgacgacacctggacaacctgctggcccagatcggcgaccagtagccgacctgtttctg ggcgaagaacctgtccgacgcatcctgtgagcgacatcctgagagtgaacaccgagatcaccaaggc ccccctgagcgcctctatgatcaagagatacagcagcaccaccaggacctgacctgtgaaagctctcgt gcggcagcagctgcctgagaagtacaaagagattttctgaccagagcaagaacggctacgcccggctacat tgacggcgaggacccagccagggaagagttctacaagttcatcaagccatcctggaaaagatggacggcaccg aggaaactgctcgtgaagctgaacagagaggacctgctgcgggaagcagcgaccttcgacaacggcagcat ccccaccagatccacctgggagagctgcacgccattctgcggcgaggaagattttaccattcctgaag gacaacgggaaaagatcgagaagatcctgaccttcgcatccctactacgtggccctctggccagggg aaacagcagattcgctggatgaccagaaagagcgaggaaaccatccccctggaacttcgaggaagtgg tggacaaggcgcttcgccagagcttcatcagcggatgaccaacttcgataagaacctgcccacagaga aggtgctgcccaagcacagcctgctgtacgagtagtaccgtgtataacgagctgaccaagtgaatactgt gaccgagggaatgagaaagcccgcttctgagcggcgagcagaaaaaggccatcgtggacctgctgttca agaccaaccggaaaagtgacctgaaagcagctgaagagaggactactcaagaaaatcgagtgttcgactccg tggaaatctccggcgtggaagatcggttcaacgcctccctgggcacataccacgatctgctgaaaattatcaag gacaaggacttctggacaatgaggaaaacgaggacattctggaagatatcgtgctgacctgacactgtttg aggacagagagatgatcaggaacggctgaaaacctatgccacctgttcgacgacaaaagtgatgaagcag ctgaagcggcgagatacaccggctggggcaggctgagccggaagctgatcaacggcatccgggacaag cagtcgggcaagacaatcctggaattctgaagtccgacggcttcgccaacagaaacttcatgagctgatcc acgacgacagcctgacctttaaagaggacatccagaagcccaggtgtccggccagggcgatagcctgcac gagcacattgccaatctggcggcagccccgccattaagaaggcgatcctgcagacagtgaaagtggtgga cgagctcgtgaaaagtgtggccggcacaagccgagaacatcgtgatcgaatggccagagagaaccag accaccagaaggacagaagaacagcggcgagagaatgaagcgatcgaagaggcgatcaagagct gggcagccagatcctgaaagaacaccccggtggaacacccagctgcagaacgagaagctgtacctgtact acctgcagaatggcgggatgtacgtggaccaggaactggacatcaaccggctgtccgactacgatgtgg acgctatcgtcctcagagctttctgaaggacgactccatcgacaacaagggtgctgaccagaagcgacaaga accggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaactactggcgagct gctgaacgccaagctgattacccagagaagttcgacaatctgaccaaggccgagagaggcgccctgagc gaactggataaggccggttcatcaagagacagctggtggaacccggcagatcacaagcacgtggcaca gatcctggactcccgatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaagtgtacac cctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaagtgcgcgagatcaacaactacc accacgcccacgacgctacctaagcgcgtggtgggaaccccgatcaaaaagtaccctaagctggaa agcgagttcgttacggcgactacaaggtgtacgacgtgcgggaagatgatcgcaagagcgagcaggaaat cggcaaggctaccgcaagtacttctacagcaacatcatgaactttttcaagaccgagattaccctggccaa cggcgagatccggaagcggcctctgatcgagacaaacggcgaaacccgggagatcgtgtgggataaggg ccgggattttgccaccgtcggaaggtgctgagcatccccaagtgaatatcgtgaaaaagaccgaggtgca gacagcggttcagcaaaaggtctatcctgccaagaggaacagcgataagctgatcgccagaaagaagg actgggaccctaagaagtacggcggttcgacagccccaccgtggcctattctgtcgtggtggccaaagt	133

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	ggaaaagggcaagtccaagaaactgaagagtgtgaaagagctgctggggatcaccatcatggaaagaagc agcttcgagaagaatcccatcgactttctggaagccaagggctacaaagaagtgaaaaaggacctgatcatca agctgcctaagtactccctgttcgagctggaaaacggccggaagagaatgctggcctctgccggcgaactgc agaagggaaacgaactggccctgccctccaaatatgtgaacttctgtacctggccagccactatgagaagct gaagggctccccgaggataatgagcagaacagctgtttgtggaacagcacaagcactacctggacgaga tcacgagcagatcagcgagttctcaagagagtatcctggccgacgctaacttgacaaagtgtgtccgc ctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccacctgtttacctgaccaat ctgggagccctgccgcttcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccaaaga ggtgtctggacgccaccctgatccaccagagcatcaccggcctgtacgagacacggatcgacctgtctcagct gggaggtgacggctcaaaaagaaccgacggcagcggaattcgagcccaagaagaagaggaaagtcG GAGGAGGAGGCAGTGGTGGGCGACTTGA AATTAGAGCCGCGTTCC TGCGCCAGAGGAATACGGCTCTCCGCACGGAGGTAGCCGA ACTTG AGCAAGAAGTACAGAGATTGGAGAACGAGGTTTCACAGTATGAG ACACGATATGGCCCCCTTGCGGCGGAAAGtaa	
NLS1- hFEN1- linker1- nCas9- linker2- T4LIG- NLS2	KRTADGSEFESPKKKRKVMGIQGLAKLIADVAPSAIRENDIKSYFGRK VAIDASMSIYQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMMEN GIKPVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGAEQEV EKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASCAALVKA GKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQEL GLNQEQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRLDP NKYPVPENWLHKEAHQLFLEPEVLDPESELKWSEPNEEELIKFMCG EKQFSEERIRSGVKRLSKSRQGSTQGRLDFFKVTGSLSSAKRKEPEP KGSTKKKAKTGAAGKFKRGKSGGSSGGSSGSETPGTSESATPSSGG SSGSSMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRH SIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYH LRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKL FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEK KNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLA QIGDQYADLFLAAKNLSAILLSDILRVNTEITKAPLSASMIKRYDEH HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR RQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRF AWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTV YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI LEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGR SRKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQK AQVSGQGDSLHEHIANLAGSPAIIKGILQTVKVVDLVKVMGRHKPE NIVIAMARENQTTQKGQKNSRERMKRIEIGIKELGSQILKEHPVENTQ LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSI DNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKS LVSDFRKDFQFYK VREINNYHHAHDAYLNA VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFF YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDF LEAKGYKEVKKDLIIKLPKYSLELENRKRMLASAGELQKGNELAL	134

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	PSKYVNFYLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF SKRVILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQLGGDSGGSS GGSKRTADGSEFESPKKKRKVSGGSSGGSMILKILNEIASIGSTKQKQ AILEKNKDNELLKR VYRLTYSRGLQYYIKKWP KPGIATQSF GMLTLT DMLDFIEFTLATRKL TGNA AIEEL TGYITD GKKDDVEVLRRVMMRD L ECGASVSIANKVWPGLIPEQPQMLASSYDEKGINKNIKFP AFAQLKAD GARCF AEVRGDELDDVRLLSRAGNEYLGLDLLKEELIKMTAEARQIH PEGVLIDGELVYHEQVKKEPEGLDFLFDAYPENSKAKEFAEVAESRT ASNGIANKSLKGTISEKEAQCMKFQVWDYVPLVEIYSLPAFRLKYDV RFSKLEQMTSGYDKVILIENQVNNLDEAKVIYKKYIDQGLEGHILKNI DGLWENARSKNLYKFKEVIDVDL KIVGIYPHRKDPTKAGGFILESECG KIKVNAGSGLKDKAGVKSHELDRTIMENQNYIYGKILECECNGWL KSDGR TDYVKLFLPIAIRLREDKTKANTFEDVFGDFHEVTGLSGGSKR TADSQHSTPPKTKRKVEFEPKKRKV	
NLS1- hFEN1- linker1- T4LIG- linker2- nCas9- NLS2	KRTADGSEFESPKKKRKVMILKILNEIASIGSTKQKQAILEKNKDNELL KR VYRLTYSRGLQYYIKKWP KPGIATQSF GMLTLT DMLDFIEFTLAT RKL TGNA AIEEL TGYITD GKKDDVEVLRRVMMRDLECGASVSIANK VWPGLIPEQPQMLASSYDEKGINKNIKFP AFAQLKADGARCF AEVRG DELDDVRLLSRAGNEYLGLDLLKEELIKMTAEARQIHPEGVLIDGELV YHEQVKKEPEGLDFLFDAYPENSKAKEFAEVAESRTASNGIANKSLK GTISEKEAQCMKFQVWDYVPLVEIYSLPAFRLKYDVRFSKLEQMTSG YDKVILIENQVNNLDEAKVIYKKYIDQGLEGHILKNIDGLWENARSK NLYKFKEVIDVDL KIVGIYPHRKDPTKAGGFILESECGKIKVNAGSGL KDKAGVKSHELDRTIMENQNYIYGKILECECNGWLKSDGR TDYVK LFLPIAIRLREDKTKANTFEDVFGDFHEVTGLSGGSSGGSSGSETPGTS ESATPESSGGSSGGSSMGIQGLAKLIADVAPSAIRENDIKSYFGRKVAI DASMSIYQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMMENGIK PVIYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGAEQEVEKF TKRLVKVTQHNDECKHLLSLMGIPYLDAPSEAEASCAALVKAGKV YAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLN QEQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHK SIEEIVRRLDPNKY PVPENWLHKEAHQLFLEPEVLDPESELK WSEPNEEELIKFMCGEKQ FSEERIRSGVKRLSKSRQGSTQGRLDDFFKVTGSLSSAKRKEPEPKGS TKKKAKTGAAGKFKRGKSGGSSGGSKRTADGSEFESPKKKRKVSGG SSGGSMDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEM AKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI QLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE DIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSR	135

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	KLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDN KVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMP QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT TVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAK GYKEVKKDLIILPKYSLFELENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSTGLYETRIDLSQLGGDSGGSKRTADS QHSTPPKTKRKVEFEPKKRKV	
NLS1- nCas9- linker1- hFEN1- linker2- T4LIG- NLS2	KRTADGSEFESPKKKRKVMDDKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAK VDDSFHRLSEESFLVEEDKKHERHPIFGNIVDE VAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEG DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNS RFAWMTRKSEETITPWNFEFVVDKGASQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQL IHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVV DELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREIN NYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIK SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGI TIMERSSEFEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQ HKHYLDEIIEQISEFSKRVI LADANLDKVL SAYNKHDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSTGLYETR IDLSQLGGDSGGSSGGSSGSETPGTSESATPSSGGSSGGSSMGIQGLA KLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQN EGETTSHLMGMFYRTIRMMENGIPVYVFDGKPPQLKSGELAKRSE RRAEAEKQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDCKHLLSL	136

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	MGIPYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLNQEQFVDLCILLGSDYCESIRGIGPKRAVDLIQKHKSIEEIVRRDPNKPYPENWLHKEAHQLFLEPEVLDPEVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRLDDEFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFGRGKSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSMILKILNEIASIGSTKQKQAILEKNKDNELLKR VYRLTYSRGLQYYIKKWP KPGIATQSFGMLTLTDMDFIEFTLATRKL TGNA AIEELTGYITDGGKDDVEVLRVMRDLCEGASVSIANKVWPGLIPEQPQMLASSYDEKGINKNIKFP AFAQLKADGARCFAEVRGDELDDVRLLSRAGNEYLGDLLEELIKMTAEARQIHPEGVLIDGELVYHEQVKKEPEGLDFLFDAYPENS KAKEFAEVAESRTASNGIANKSLKGTISEKEAQCMKFQVWDYVPLVEIYSLPAFRLKYDVRFSKLEQMTSGYDKVILIENQV VNNLDEAKVIYKKYIDQGLEG IILKNIDGLWENARSKNLYKFKEVIDVDLKIVGIYPHRKDPTKAGGFILESECGKIKVNAGSGLKDKAGVKSHELDRTIMENQNYIIGKILECECNGWLKSDGR TDYVKLFLPIAIRLREDKTKANTFEDVFGDFHEVTGLSGGSKRTADSQHSTPPKTKRKVEFEPKKRKV	
NLS1-T4LIG-linker1-nCas9-linker2-hFEN1-NLS2	KRTADGSEFESPKKKRKVMILKILNEIASIGSTKQKQAILEKNKDNELLKR VYRLTYSRGLQYYIKKWP KPGIATQSFGMLTLTDMDFIEFTLATRKL TGNA AIEELTGYITDGGKDDVEVLRVMRDLCEGASVSIANKVWPGLIPEQPQMLASSYDEKGINKNIKFP AFAQLKADGARCFAEVRGDELDDVRLLSRAGNEYLGDLLEELIKMTAEARQIHPEGVLIDGELVYHEQVKKEPEGLDFLFDAYPENS KAKEFAEVAESRTASNGIANKSLKGTISEKEAQCMKFQVWDYVPLVEIYSLPAFRLKYDVRFSKLEQMTSGYDKVILIENQV VNNLDEAKVIYKKYIDQGLEG IILKNIDGLWENARSKNLYKFKEVIDVDLKIVGIYPHRKDPTKAGGFILESECGKIKVNAGSGLKDKAGVKSHELDRTIMENQNYIIGKILECECNGWLKSDGR TDYVKLFLPIAIRLREDKTKANTFEDVFGDFHEVTGLSGGSSGGSSGSETPGTSESATPESGGSSGGSSMDKKYSIGLDIGTNSVGWAVITDEYKVP SKKF KVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKD TYD DDLNLLAQIGDQYADLFLAAKNLS DAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA SQEEFYKF I KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKH SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGT YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA I KKGILQTVKVVD ELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIE EG I KELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIV PQSFLKDDSIDNKVLTRSDKNRGKSDNPSEE VVKMKMKNYWRQLLN AKLITQRKFDNLTKAERGG LSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYK VREINNYHH	137

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	AHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG RDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKEKVKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENGRKRMLASAG ELQKGNELALPSKYVNFLYLASHYEKLKSGPEDNEQKQLFVEQHKH YLDEIIEQISEFSKRVLADANLDKVL SAYNKHHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYETRIDL S QLGGDSGGSSGGSKRTADGSEFESPKKKRKVSGGSSGGSGMGIQGLAK LIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQNE EGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKRSE RAEAEKQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDCKHLLSLM GIPYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHL TASEAKKLPIQEFHLSRILQELGLNQE QFVDLCILLGSDYCESIRGIGPK RAVDLIQKHKSIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPEVLDPE SVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSGTQGR L DDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKRGKSGGSKRT ADSQHSTPPKTKRKVEFEPKKRKV	
NLS1- nCas9- linker1- T4LIG- linker2- hFEN1- NLS2	KRTADGSEFESPKKKRKVMDDKYSIGLDIGTNSVGWAVITDEYKVPS KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDE VAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNLGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAP LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEFVVDK GASAQSFIERMTNFDKNLPNEKV LPKHSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFML IHDDSLTFKEDIQKAQVSGQGDLSLHEHIANLAGSPAICKGILQTVKVV DELVKVMGRHKPENIVIEAMARENQTTQKGQKNSRERMKRIEIGIKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVRIN NYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIK SEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKEKVKSKKLKSVKELLGI TIMERSSEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENGRKRML ASAGELQKGNELALPSKYVNFLYLASHYEKLKSGPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHHRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYETR IDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMILKILNE IASIGSTKQKQAILEKNKDNELLKRVYRLTYSRGLQYYIKKWPKPGIA	138

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	TQSF GMLTLTDM LDFIEFTLATRKL TGNA AIEELTGYITD GKKDDVEV LRRVMMRDLECGASVSIANKVWPGLIPEQPQMLASSYDEKGINKNIK FPFAAQLKADGARCFAEVRGDELDDVRLLSRAGNEYLGDLLEELI KMTAEARQIHPEGVLIDGELVYHEQVKKEPEGLDFLFDAYPENSKAK EFAEVAESRTASNGIANKSLKGTISEKEAQCMKFQVWDYVPLVEIYS LPAFRLKYDVRFSKLEQMTSGYDKVILIENQVVNNLDEAKVIYKKYI DQGLEGILKNIDGLWENARSKNLYKFKEVIDVDLKIVGIYPHRKDPT KAGGFILESECGKIKVNAGSGLKDKAGVKSHELDRTIMENQNYIIG KILECECNGWLKSDGR TDYVKLFLPIAIRLREDKTKANTFEDVFGDFH EVTGLSGGSSGGSKRTADGSEFESPKKKRKVS GGSSGGSMGIQGLAK LIADVAPSAIRENDIKSYFGRKVAIDASMSIYQFLIAVRQGGDVLQNE EGETTSHLMGMFYRTIRMMENGIKPVYVFDGKPPQLKSGELAKR SER RAEAEKQLQQAQAAGAEQEVEKFTKRLVKVTKQHNDCKHLLSLM GIPYLDAPSEAEASCAALVKAGKVYAAATEDMDCLTFGSPVLMRHL TASEAKKLPIQEFHLSRILQELGLNQE QFVDLCILLGSDYCESIRGIGPK RAVDLIQKHK SIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPEVLDPE SVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLSKSRQGSTQGRL DDFFKVTGSLSSAKRKEPEPKGSTKKKAKTGAAGKFKRGKSGGSKRT ADSQHSTPPKTKRKVEFEPKKRKV	
NLS1- T4LIG- linker1- hFEN1- linker2- nCas9- NLS2	KRTADGSEFESPKKKRKVMILKILNEIASIGSTKQKQAI LEKNKDNE LL KR VYRLTYSRGLQYYIKKWP KPGIATQSF GMLTLTDM LDFIEFTLAT RKL TGNA AIEELTGYITD GKKDDVEV LRRVMMRDLECGASVSIANK VWPGLIPEQPQMLASSYDEKGINKNIKFPFAAQLKADGARCFAEVRG DELDDVRLLSRAGNEYLGDLLEELIKMTAEARQIHPEGVLIDGELV YHEQVKKEPEGLDFLFDAYPENSKAKEFAEVAESRTASNGIANKSLK GTISEKEAQCMKFQVWDYVPLVEIYSLPAFRLKYDVRFSKLEQMTSG YDKVILIENQVVNNLDEAKVIYKKYIDQGLEGILKNIDGLWENARSK NLYKFKEVIDVDLKIVGIYPHRKDPTKAGGFILESECGKIKVNAGSGL KDKAGVKSHELDRTIMENQNYIIGKILECECNGWLKSDGR TDYVK LFLPIAIRLREDKTKANTFEDVFGDFHEVTGLSGGSSGGSSGSETPGTS ESATPESSGGSSGGSSSMGIQGLAKLIADVAPSAIRENDIKSYFGRKVAI DASMSIYQFLIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMMENGIK PVYVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGAEQEVEKF TKRLVKVTKQHNDCKHLLSLMGIPYLDAPSEAEASCAALVKAGKV YAAATEDMDCLTFGSPVLMRHLTASEAKKLPIQEFHLSRILQELGLN QE QFVDLCILLGSDYCESIRGIGPKRAVDLIQKHK SIEEIVRRLDPNKY PVPENWLHKEAHQLFLEPEVLDPE SVELKWSEPNEEELIKFMCGEKQ FSEERIRSGVKRLSKSRQGSTQGRLDDFFKVTGSLSSAKRKEPEPKGS TKKKAKTGAAGKFKRGKSGGSSGGSKRTADGSEFESPKKKRKVS GG SSGGSMDKKYSIGLDIGTNSVGWAVITDEYK VPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE M AKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFI QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQI GDQYADLFLAAKNLSAILLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKP ILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP	139

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDN KVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEND KLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMP QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS TVAYSVLVVAKEKVKSKKLKSVKELLGITIMERSSEKKNPIDFLEAK GYKEVKKDLIKLPKYSLENGRKRMLASAGELQKGNELALPSKY VNFLYLASHYEKLKGPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRV LADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDT TIDRKRYTSTKEVLDATLIHQSTGLYETRIDLSQLGGDSGGSKRTADS QHSTPPKTKRKVEFEPKKKKRV	
NLS1-T5 EXO- linker1- nCas9- linker2- T4LIG- NLS2	KRTADGSEFESPKKKRKVMSKSWGKFIEEEEAEMASRRNLMIVDGT NLGFRFKHNSKPPFASSVSTIQSLAKSYSARTTIVLGDKGKSVFRL EHLPEYKGNRDEKYAQRTEEEKALDEQFFEYLKDAFELCKTTFPTFTI RGVEADDMAAYIVKLIGHLYDHVWLISTDGDWDTLTLDKVSRSFST TRREYHLRDMYEHNVDDVEQFISLKAIMGDLGDNIRGVEGIGAKR GYNIREFGNVLDIIDQLPLPGKQKYIQNLNASEELLFRNLILVDLPTYC VDAIAAVGQDVLDFKTDILEIAEQSGGSSGGSSGSETPGTSESATPES SGGSSGGSSMDKKYSIGLDIGTNSVGWAVITDEYKVPKSKFKVLGNT DRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLV DSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLNLAQLPG EKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNL LAQIGDQYADLFLAAKNLSAAILSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYK FIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYFPLKDNREKIEKILTRIPYVVGPLARGNSRFAWMTRKSEE TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENE DILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLRSKLINGIRDKQSGKTILDFLKSDFANRNFQMQLIHDDSLTFKEDIQ KAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHK PENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKD DSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQR KFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTK YDENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYL NAVVG TALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAK	140

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR KVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSSFENPI DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS EFSKRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTL TNLGAPA AFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDSGG SSGGSKRTADGSEFESPKKKRKVSGGSSGGSMILKILNEIASIGSTKQK QAILEKNKDNELLKR VYRLTYSRGLQYYIKKWP KPGIATQSF GMLTL TDM LDFIEFTLATRKL TGNA AIEEL TGYITDGKKDDVEVLRRVMMRD LECGASVSIANKVWPGLIPEQPQMLASSYDEKGINKNIKFP AFAQLKA DGARCF AEVRGDELDDVRLLSRAGNEYLGLDLLKEELIKMTAEARQI HPEGVLIDGELVYHEQVKKEPEGLDFLDAYPENSKAKEFAEVAESR TASNGIANKSLKGTISEKEA QCMKFQVWDYVPLVEIYSLPAFRLKYD VRFSKLEQMTSGYDKVIL IENQV VNNLDEAKVIYKKYIDQGLEG IILK NIDGLWENARSKNLYKFKEVIDVDL KIVGIYPHRKDPTKAGGFILESE CGKIKVNAGSGLKDKAGVKSHELDRTRIMENQNYI GKILECECNG WLKSDGR TDYVKLFLPIA IRLREDKTKANTFEDVFGDFHEVTGLSGG SKRTADSQHSTPPKTKRKVEFEPK KKRKV	
LZ-SplintR-bpNLS	ATGTTGGAGATAGAGGCTGCTTTCCTTGAACGGGAGAATACGGCC CTCGAGACTAGGGTTGCTGAGCTTAGACAGCGAGTCCAAAGACTG CGAAACCGGGTGTCCCAATATAGGACCAGATACGGACCTCTGGGT GGAGGGAAATCCGGTGGGAGTAGCGGCGGGTCTAGTGGCTCAGA GACACCTGGCACGAGCGAGAGTGCGACTCCTGAAAGCTCCGGCGG CAGCAGTGGGGGAAGTTCCATGGCAATCACTAAGCCCCTCTTGGC GGCGACTTTGGAAAACATCGAGGATGTGCAATTCCCGTGCCTTGC CACACCAAAGATAGACGGGATCCGATCAGTGAAGCAAACGCAGA TGCTCTCTAGAACGTTCAAGCCTATTAGAACTCAGTGATGAATC GGCTCTTGACTGAGCTGTTGCCGGAAGGCAGCGATGGGGAAATAT CTATCGAGGGAGCCACATTTCAAGACACTACGAGCGCCGTAATGA CTGGACATAAGATGTATAATGCTAAATTCTCCTACTATTGGTTTGA CTATGTCACTGATGACCCTCTTAAGAAATATATAGACCGAGTGGA GGATATGAAAAATTATATTACTGTACACCCGCATATTCTGGAACA TGCCCAAGTTAAGATTATTCCTCTCATTCCCGTCGAGATTAATAAT ATCACAGAACTGCTTCAGTATGAGCGCGACGTATTGTCCAAAGGC TTTGAAGGGGTATGATACGCAAACCGGACGGCAAGTACAAGTTC GGAAGAAGCACATTGAAAGAGGGTATATTGCTGAAGATGAAGCA GTTTAAGGATGCTGAGGCAACAATAATCAGCATGACAGCACTTTT TAAAAATACCAACACGAAAATAAGGACAATTTTGGTTATAGTAA GCGGTCAACGCACAAAAGTGGAAGGTAGAAGAAGACGTAATGG GTAGCATTGAGGTGGATTATGACGGGGTGGTTTTTCAGCATAGGGA CTGGGTTTGATGCAGATCAACGGAGGGACTTTTGGCAGAACAAAG AATCATATATAGGCAAATGGTAAAGTTCAAATACTTCGAAATGG GAAGTAAAGACTGCCCCAGATTCCCTGTATTCATTGGCATCAGGC ACGAGGAGGACAGGAGTGGGGGATCAAAGCGGACTGCTGATAGT CAGCATAGTACTCCACCAAGACCAAGCGGAAAGTTGAGTTTGAG CCGAAGAAAAAGCGAAAAGTGTA	141

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
LZ-T4LIG-bpNLS	ATGCTTGAGATCGAGGCGGCGTTCCTCGAAAGAGAGAACACCGCA CTTGAAACTCGCGTGCGCAGAATTGAGGCAGCGGGTGCAAAGACTT AGAAATAGAGTCTCTCAGTATCGGACCCGATATGGTCCTCTGGGA GGCGGGAAGTCTGGAGGTTCAAGCGGAGGCAGTTCCGGGAGTGA GACACCGGGAACCTTCTGAGAGTGCAACTCCTGAGAGCTCTGGTGG ATCATCCGGAGGCTCCAGTATGATTCTTAAATTTCTTAACGAGATT GCGAGTATTGGCAGCACGAAACAAAAGCAGGCCATATTGGAAAA GAATAAGGACAATGAGTTGCTTAAACGCGTGTATAGGCTCACTTA CTCTCGCGGACTGCAATACTATATTAAAAAATGGCCTAAGCCCGG CATCGCTACTCAAAGCTTCGGAATGCTTACGCTGACAGATATGCTC GACTTCATCGAGTTTACTCTCGCAACAAGGAAGTTGACTGGCAAC GCCGCGATTGAAGAATTGACGGGTTATATCACGGACGGGAAGAA GGATGATGTTGAGGTGCTGAGGCGCGTATGATGCGCGACCTCGA ATGTGGTGCCTCAGTTTCCATAGCCAATAAAGTTTGGCCAGGCTTG ATCCCGGAGCAGCCACAGATGCTGGCCAGTAGCTACGACGAGAA GGGTATTAACAAAAATATCAAGTTTCCAGCGTTTGCACAACTTAA AGCGGATGGGGCGCGGTGTTTCGCCGAAGTCCGGGGTGACGAATT GGACGATGTGCGCCTTCTGAGTCGCGCAGGAAATGAATATCTGGG GCTTGACCTCTTGAAGGAGGAGCTGATTAAGATGACAGCAGAAGC CAGGCAGATCCATCCAGAGGGGGTACTTATTGATGGTGAACCTCGT ATACCATGAGCAGGTTAAGAAGGAGCCAGAGGGTTTGGATTTCCT CTTTGACGCCTATCCCGAGAATTCAAAGGCAAAGGAGTTCGCCGA GGTTGCAGAAATCAAGAACGGCTTCCAACGGCATAGCGAATAAATC ACTCAAAGGAACTATATCTGAAAAGGAGGCACAGTGTATGAAATT CCAAGTGTGGGACTATGTGCCGCTTGTTCGAGATTTACAGCTTGCTT GCTTTCCGATTGAAGTACGATGTACGGTTTAGTAAGCTCGAGCAA ATGACTTCAGGTTACGATAAAGTCATCTTGATTGAGAACCAGGTC GTTAATAATCTTGACGAGGCGAAGGTCATATATAAGAAATATATA GATCAAGGGCTCGAGGGTATCATTCTGAAGAATATAGATGGCTTG TGGGAAAACGCCAGGTCCAAAAACCTGTATAAGTTTAAGGAAGTA ATAGATGTAGATTTGAAAATAGTTTGAATTTACCCCATCGGAAG GACCCACGAAAGCGGGTGGGTTTATCCTCGAGAGCGAGTGTGGG AAGATAAAAGTGAATGCCGGCTCCGGATTGAAGGACAAGGCAGG TGTGAAAAGTCATGAGCTCGATCGGACGAGAATAATGGAAAACC AGAATTACTACATTGGAAAGATTTTGGAAATGCGAGTGTAAACGGCT GGTTGAAGAGCGACGGACGCACCGATTACGTGAAACTCTTTCTGC CAATTGCAATCAGGTTGAGAGAGGATAAGACTAAGGCCAATACTT TCGAGGACGTCTTCGGAGACTTTTACGAAGTCACTGGGCTTTCTGG GGGTAGTAAGAGAACTGCAGATAGCCAGCATTCAACGCCGCCAA AAACAAAGCGAAAGGTAGAATTCGAACCAAAGAAAAAGCGGAAA GTATAA	142
LZ-hLIG1(233-919)-bpNLS	ATGCTCGAGATCGAAGCTGCATTTCTGGAGAGGGAGAATACCGCC CTCGAAACCCGGGTGGCTGAATTGCGACAGAGAGTGCAACGGCTC CGGAATAGAGTATCTCAATATCGAACCCGCTATGGGCCTCTCGGA GGGGGTAAATCTGGCGGAAGTTCTGGCGGTAGTTCAGGAAGTGAG ACACCGGGAAGTGTGAATCCGCGACTCCCGAATCAAGTGGGGGA TCATCTGGAGGGTCAAGCACACCCAGGAAACCAGCCGTGAAAAA AGAGGTTAAAGAAGAGGAACCTGGGGCTCCGGGAAAGGAGGGAG CAGCGGAAGGTCCGCTCGACCCTTCAGGATACAACCCAGCCAAAA	143

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	ACAACTACCACCCCGTAGAGGATGCTTGCTGGAAGCCAGGCCAAA AGGTGCCCTATTTGGCCGTTGCTAGGACTTTCGAAAAAATTGAGG AGGTGAGCGCGCGACTCAGAATGGTAGAGACTCTGTCTAACCTCC TTCGCTCCGTAGTGGCTCTTTCACCTCCAGATCTTCTTCCAGTGCTG TACCTGAGCCTGAACCACTTGGGCCCTCCCCAGCAGGGACTGGAA CTGGGCGTAGGGGACGGAGTATTGCTGAAGGCTGTTGCTCAGGCA ACCGGACGACAGCTCGAGTCTGTGCGAGCAGAAGCTGCAGAAAA GGGGGACGTCGGGTTGGTTGCCGAAAATTCAAGATCTACCCAACG ATTGATGTTGCCACCGCCGCCTCTGACTGCGTCAGGTGTATTCTCC AAGTTCCGGGATATTGCCAGGCTTACGGGTAGCGCTTCCACTGCT AAAAAGATCGACATAATAAAAGGTCTGTTCGTCGCTTGTCGCCAT TCAGAGGCGAGGTTTATAGCCAGATCCCTTTCGGGACGACTTCGA CTCGGCTTGGCTGAGCAGTCAGTACTGGCAGCTTTGTCTCAAGCTG TATCACTCACGCCCCCGGACAAGAATTTCCACCCGCCATGGTTG ACGCAGGCAAGGGTAAGACTGCTGAGGCAAGAAAGACGTGGCTG GAGGAACAAGGTATGATACTTAAACAAACGTTTTGCGAAGTTCCG GACTTGGACCGGATCATACCTGTGTTGCTGGAGCACGGCCTCGAG CGCTTGGCCGAACACTGTAACTGTCTCCAGGAATACCTCTCAA CCCATGTTGGCTCATCTACGAGGGGAATCTCAGAGGTACTTAAA CGGTTTGAAGAAGCCGCTTTCACGTGCGAATACAAGTATGATGGT CAGAGAGCGCAAATCCACGCATTGGAAGGGGGTGAGGTAAAGAT TTTTCAAGGAATCAGGAGGACAATACAGGGAAGTACCCCGATAT CATCAGTCGGATTCTTAAATTAAGCTTCCATCAGTCACGTCCTTC ATACTGGACACTGAGGCAGTGGCTTGGGACCGAGAGAAGAAGCA GATACAACCCTTTCAGGTACTTACAACCAGAAAGCGCAAGGAAGT CGACGCTTCTGAGATTCAAGTACAAGTCTGCCTTTATGCGTTTGAC CTGATCTATCTTAATGGAGAGAGTTTGGTGAGAGAACCCTTGAGC AGACGACGGCAGCTCTTGAGAGAAAAATTCGTAGAACTGAGGG GGAGTTCGTCTTTGCGACTAGTCTCGACACCAAAGACATTGAGCA AATCGCGGAATTCTCGAACAGTCAGTTAAAGACTCCTGCGAAGG TCTGATGGTTAAGACTCTTGACGTGGATGCTACCTACGAGATAGCT AAGCGGTCACACAATTGGCTGAACTGAAAAAGGACTATCTGGAT GGAGTTGGGGACACGCTGGATTTGGTCGTTATCGGGGCCTATCTG GGACGCGGTAAGCGGGCAGGGAGATATGGTGGATTCTCCTCGCT TCATACGATGAGGACTCTGAAGAGCTGCAGGCTATATGCAAACCT GGGACGGGTTTTTCCGATGAAGAATTGGAGGAACATCATCAGTCA CTGAAGGCCCTTGTATTGCCAAGTCCACGCCCATACGTACGAATC GATGGAGCAGTAATCCCTGACCACTGGCTTGACCCGTCCGCCGTC TGGGAAGTAAAGTGCGCGGATCTCTCTCAGTCCGATCTACCCA GCCGCACGGGGGCTGGTTGACAGTGACAAGGGTATCAGCCTGCGA TTCTCTGATTATACGCGTCCGGGAAGACAAGCAACCGGAACAG GCTACGACCTCTGCACAGGTGCGATGTTTGTATAGAAAACAGAGC CAAATTCAGAATCAACAAGGCGAAGACAGTGGGTCCGATCCTGAA GATACCTACTCAGGCGGCAGTAAACGGACAGCTGATAGCCAACAC TCAACTCCTCCGAAGACTAAAAGGAAGGTAGAGTTTGAACCAAAA AAGAAAAGGAAAGTGTA	
LZ-hLIG1(1	ATGCTCGAGATCGAGGCGGCGTTTCCTTGAACGCGAGAACACTGCG CTGGAAACGAGGGTCGCGGAACCTCCGCCAGAGGGTTCAACGGTTG AGGAATCGAGTGAGTCAGTACCGAACCCGATATGGACCACTGGGT	144

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
19-919)- bpNLS	GGCGGGAAATCAGGGGGGCTCATCCGGCGGCTCCAGCGGGAGCGA AACCCCGGGTACCTCAGAATCTGCGACGCCAGAAAGCTCAGGCGG ATCTAGCGGCGGTAGTTCACCGAAGCGCCGGACTGCACGAAAGCA ACTGCCAAAACGGACTATACAAGAAGTCCTGGAAGAACAAAGCG AAGATGAGGATCGCGAAGCCAAGCGCAAGAAAGAGGAAGAGGA AGAAGAGACTCCAAAGGAGTCCTTGACCGAAGCAGAAGTCGCAA CGGAGAAGGAAGGTGAGGATGGGGATCAGCCAACAACCCCGCCT AAACCTCTGAAAACCTCTAAGGCGGAGACACCAACTGAGAGTGTC AGCGAACCGGAGGTAGCCACGAAACAAGAGCTTCAGGAGGAAGA AGAACAGACAAAGCCACCTCGGCGGGCTCCCAAAACCTTAGCTC CTTCTTCACGCCTCGAAAGCCAGCAGTGAAGAAAGAAAGTGAAGGA GGAGGAACCTGGCGCCCCCTGGAAAGGAGGGCGCAGCCGAGGGCC CGCTGGACCCCTCAGGGTATAACCCGGCAAAAAATAATTACCACC CGGTCGAGGACGCTTGTTGGAAACCAGGCCAAAAGGTACCTTACC TCGCCGTCGCTAGGACCTTTGAGAAGATAGAGGAAGTTAGTGCTA GGTTGAGAATGGTCGAAACCCTTAGTAACCTTCTCAGGTCCGTAG TCGCCCTTAGTCCCCCAGACCTGCTTCCGGTGCTGTACCTGTCCCT GAACCATCTCGGTCCCCCCCCAACAGGGACTGGAGTTGGGCGTCGG TGACGGCGTTCTCCTGAAAGCGGTTGCACAAGCTACAGGAAGGCA ACTGGAATCTGTCCGGGCTGAGGCTGCAGAGAAAGGTGACGTGGG GCTTGTTGGCAGAGAATAGTCGGTCAACACAGCGGCTGATGCTGCC ACCGCCCCCGCTTACGGCTAGTGGGGTATTCTCCAAATTTAGAGAT ATAGCACGGCTGACGGGATCAGCTTCCACTGCGAAGAAGATCGAT ATCATTAAGGGTTTGTTCGTGGCTTGCAGGCATTCCGAAGCACGCT TCATTGCACGCTCCCTTTCAGGGAGACTCAGACTTGGGCTGGCCG AGCAATCTGTACTGGCGGCCCTGTCTCAGGCGGTGAGCCTTACGC CGCCCGGGCAAGAGTTCCCTCCTGCGATGGTCGATGCTGGGAAGG GAAAAACCGCCGAAGCTCGAAAAACATGGCTGGAGGAGCAAGGA ATGATTTTGAAGCAGACGTTCTGTGAAGTACCGGACTTGGATCGC ATCATACCTGTGCTTCTCGAACATGGTTTGGAGCGGCTCCCCGAGC ATTGCAAACCTCTCTCCGGGCATCCCCCTCAAGCCAATGCTCGCCCA CCCCACGCGCGGAATCAGTGAGGTACTGAAACGCTTTGAAGAGGC AGCGTTTACTTGTGAATACAAGTACGATGGCCAAAGGGCACAAAT TCATGCACTTGAAGGCGGGGAAGTTAAGATATTCAGCAGGAATCA GGAGGACAACACGGGAAAATATCCTGACATAATATCTCGAATCCC TAAATTAAGTTGCCTAGCGTAACCAGCTTCATCCTGGATACCGA AGCCGTGGCGTGGGATAGGGAGAAAAAGCAAATACAGCCATTTC AGGTGCTTACAAC TAGAAAACGAAAAGAGGTGGACGCTAGTGAA ATCCAAGTCCAGGTATGTCTTTATGCCTTCGATTGATATACCTTA ATGGTGAGTCCCTTGTACGGGAACCGCTTAGTAGGAGGCGGCAGT TGCTGAGGGAAAATTTTGTGAAACTGAGGGAGAGTTTGTATTTG CAACGTCATTGGATACAAAGGACATAGAACAATAGCAGAATTTTC TGGAGCAGTCAGTAAAAGACTCCTGCGAGGGCCTGATGGTGAAAA CTCTTGATGTGGACGCCACTTATGAAATCGCAAAAAGGTCACACA ATTGGCTGAAACTTAAAAAGGATTACTTGGACGGGGTCGGGGATA CCCTCGATCTCGTCGTAATCGGAGCTTATCTCGGTAGGGGGAAGC GAGCCGGGCGATACGGAGGCTTTCTCTTGGCTAGTTATGACGAAG ATTCCGAAGAGCTGCAGGCCATATGCAAGCTTGGAACGGGTTTCA GCGATGAGGAATTGGAGGAGCATCATCAGAGCTTGAAGGCACTG	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	GTGCTCCCCTCTCCTAGGCCGTACGTTAGAATAGACGGAGCAGTG ATACCCGATCATTGGCTCGATCCGTCAGCTGTTTGGGAGGTGAAG TGTGCAGACCTGTCCCTCTCTCCTATTTACCCTGCAGCACGCGGTC TGGTTGACTCTGACAAAGGGATTAGCTTGAGGTTCCCTAGATTTAT TCGGGTGCGCGAAGACAAACAGCCTGAACAGGCGACAACGTCCG CGCAGGTGCGCATGCCTTTATCGAAAACAGAGTCAGATCCAGAATC AACAAGGAGAAGATTCAAGGAGTGACCCGGAGGACACTTATAGT GGCGGCTCAAAACGAACCGCCGATAGTCAGCATTCAACACCTCCA AAACTAAAAGGAAAGTCGAGTTTGAGCCAAAGAAGAAGCGCAA AGTCTAA	
T4-LZ	ATGATCCTTAAGATTCTCAACGAAATCGCTAGTATAGGGTCCACT AAGCAGAAGCAGGCCATATTGGAAAAAATAAGGACAATGAACT TTTGAAGAGAGTCTATAGACTGACGTACTCTAGGGGGCTCCAGTA CTACATCAAGAAATGGCCTAAACCTGGCATTGCGACGCAGTCATT CGGTATGCTGACATTGACGGATATGTTGGATTTTCATTGAGTTTACG CTGGCCACCAGAAAACCTTACGGGTAATGCTGCGATAGAAGAACTT ACAGGGTACATAACAGACGGGAAGAAAGATGACGTGGAAGTGCT CAGACGAGTTATGATGCGCGATCTCGAGTGCGGCGCTAGCGTGTC AATCGCGAACAAGTCTGGCCCGGCCTCATACCAGAGCAGCCACA GATGCTGGCATCTTCCTATGACGAAAAAGGCATAAACAAGAATAT TAAATTCCTCGGCCTTCGCTCAACTCAAAGCAGATGGTGCCAGGTG TTTTGCCGAAGTTCGGGGTGATGAACTTGATGACGTGCGGCTCTTG TCTAGGGCAGGTAAACGAGTACCTCGGCCTGGACTTGCTTAAAGAG GAACTGATTAAAATGACAGCTGAGGCGCGGCAGATACACCCCGA GGGCGTCCTTATCGACGGGGAGCTGGTGTATCACGAACAAGTTAA AAAGGAACCGGAGGGTCTTGATTTTCTTTTCGACGCGTATCCTGAG AACAGCAAGGCGAAAGAATTTGCAGAAGTTGCAGAAAGCAGGAC CGCAAGTAATGGAATCGCTAATAAAAGCCTCAAGGGTACCATCAG CGAAAAAGAAGCCCAGTGCATGAAATTTCAAGTTTGGGACTATGT CCCCTTGGTTCGAAATTTACTCCCTGCCCGCATTCCGGCTGAAGTAT GATGTTTCGCTTCAGTAAACTGGAGCAAATGACGAGCGGTTATGAT AAGGTTATACTTATTGAGAATCAGGTCGTAAATAATTTGGACGAG GCGAAAGTTATATACAAAAAGTATATAGACCAAGGGTTGGAGGG GATCATTTTGAAGAACATAGACGGACTTTGGGAGAACGCCCGGTC CAAGAATTTGTATAAATTCAAAGAAGTCATAGATGTTGACCTCAA GATAGTAGGTATATATCCCCACAGAAAGGACCCAACCAAAGCAG GCGGATTCATTTTGGAGTCCGAGTGTGGGAAGATAAAGGTCAATG CTGGATCTGGACTCAAGGACAAAGCTGGTGTGAAGTCACATGAAC TGGACCGAACCAGGATTATGGAGAATCAGAACTATTACATCGGGA AGATATTGGAGTGTGAATGCAACGGCTGGCTTAAATCAGATGGAA GAACTGATTACGTTAAATTGTTCTGCCCATAGCCATACGACTCCG CGAGGACAAAACGAAGGCTAACACGTTTGAAGACGTATTCGGAG ATTTCCATGAGGTGACTGGCCTTAGTGgaggctccaaacggacagcagactccaa cattcaacacccccaaaaaacaagcggaaggtagagtttgagccaaaaaagaaaagaaggtcGGAGG AGGAGGCAGTGGTGGGCGACTTGAAATTAGAGCCGCGTTCTTGCG CCAGAGGAATACGGCTCTCCGCACGGAGGTAGCCGAACCTTGAGCA AGAAGTACAGAGATTGGAGAACGAGGTTTTCACAGTATGAGACAC GATATGGCCCCCTTGGCGGCGGAAAGTAA	145

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
LZ-hLIG4(1-620)	ATGCTTGAGATCGAGGCGGCGTTCCTCGAAAGAGAGAACACCGCA CTTGAAACTCGCGTGGCAGAATTGAGGCAGCGGGTGCAAAGACTT AGAAATAGAGTCTCTCAGTATCGGACCCGATATGGTCCTCTGGGA GGCGGGAAGTCTGGAGGTTCAAGCGGAGGCAGTTCGGGAGTGA GACACCGGGAACCTTCTGAGAGTGCAACTCCTGAGAGCTCTGGTGG ATCATCCGGAGGCTCCAGTGCAGCTTCTCAGACCTCTCAAACAGT AGCCTCTCATGTACCGTTCGCTGACTTGTGTTCTACGCTCGAACGC ATCCAGAAATCAAAGGGGCGCGCCGAGAAAATCCGGCACTTCAG AGAATTCCTGGATTCTGGAGGAAGTTTCATGATGCTCTCCACAAA AATCACAAAGATGTAACGGATAGTTTCTACCCTGCTATGAGACTT AACTGCCGCAGCTTGAGAGGGAACGCATGGCGTATGGTATAAAG GAGACAATGTTGGCGAAATTGTATATTGAGCTGCTGAACTTGCCA AGAGATGGAAAGGACGCGCTCAAACCTGCTGAACTATAGAACC CACGGGTACCCATGGTGACGCCGGTGACTTTGCCATGATCGCCTA TTTCGTAAGTCAAACCTCGATGTCTTCAAAAAGGTTCTCTTACAATT CAGCAAGTCAACGACCTGCTGGATTCAATTGCGAGTAACAACAGC GCTAAGCGAAAGGATCTCATTAAAGAAAAGCCTCCTGCAGCTGATA ACTCAGTCCTCTGCACTCGAACAAAAATGGCTGATTTCGGATGATT ATCAAGGATTTGAAGTTGGGGGTATCTCAGCAAACTATTTTCAGC GTGTTTCACAATGATGCAGCAGAATTGCATAATGTCACAACAGAT CTTGAGAAAGTCTGCCGACAGTTGCACGACCCCTCTGTAGGCTTG AGTGACATATCTATAACACTTTTTTCTGCGTTCAAACCCATGTTGG CTGCTATTGCGGACATAGAACACATCGAGAAAGACATGAAACATC AGTCATTCTATATAGAGACTAAATTGGACGGCGAGAGGATGCAAA TGCACAAAGATGGTGATGTGTATAAATATTTTCCCGCAACGGCT ACAACTACACTGATCAATTTCGGAGCGTCCCCAACTGAAGGGTCCC TCACTCCTTTCATACACAATGCGTTTAAAGGCCGATATTCAGATATG TATCCTCGACGGCGAAATGATGGCGTACAATCCCAATACCCAGAC CTTCATGCAAAAGGGAACGAAGTTCGATATTAAACGGATGGTTGA AGATTCCGACCTCCAAACATGTTACTGTGTGTTTGATGTCTCTGATG GTGAATAACAAAAAACTCGGCCATGAAACCCCTCGAAAGCGATAC GAAATACTCAGCAGTATATTTACTCCAATACCAGGCCGAATCGAG ATCGTACAGAAAACACAAGCCATACTAAGAATGAAGTTATTGAT GCACTGAACGAAGCCATAGACAAGAGGGAAGAAGGCATAATGGT CAAGCAGCCTCTGAGTATATATAAACCTGACAAAAGGGGGGAAG GATGGCTGAAGATAAAGCCAGAATACGTGTCTGGTCTTATGGACG AATTGGACATTCTCATCGTTCGGAGGATATTGGGGTAAGGGTTCCA GGGGGGGGATGATGTCCCACTTTCTGTGTGCGGTTGCCGAGAAAC CGCCCCAGGGGAAAAACCATCAGTGTTCATACGTTGTACACGCG TCGGCTCAGGTTGTACGATGAAGGAACCTTACGATCTGGGGTTGA AACTCGCCAAATATTGGAAGCCATTCCATCGGAAAGCACCGCCCT CTAGTATCTTGTGTGGGACGGAGAAGCCAGAAGTTTATATAGAGC CATGTAACCTCAGTAATTGTTCAAATCAAAGCCGCAGAGATCGTCC CGTCAGACATGTACAAGACTGGATGCACCCTTAGATTTCTCTCGCAT CGAAAAAATAAGAGATGATAAAGAGTGGCATGAGTGCATGACTC TTGACGACCTTGAACAGCTCCGCGGGAAGGCCAGCGGTAAACTGG CTAGTAAGCACCTCTACATCGGGGGTGACAGTGgaggtccaaacggacag cagactcccaacattcaacacccccaaaaacaaagcggaaggtagagtttgagccaaaaaagaaaagaaag gtctaa	146

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
LZ-nCas9	ATGCTTGAGATCGAGGCGGCGTTCCTCGAAAGAGAGAACACCGCA CTTGAAACTCGCGTGGCAGAATTGAGGCAGCGGGTGCAAAGACTT AGAAATAGAGTCTCTCAGTATCGGACCCGATATGGTCCTCTGGGA GGCGGGAAGTCTGGAGGTTCAAGCGGAGGCAGTTCCGGGAGTGA GACACCGGGAACCTTCTGAGAGTGCAACTCCTGAGAGCTCTGGTGG ATCATCCGGAGGCTCCAGTaaacggacagccgacggaagcaggttcgagtcaccaaga agaagcggaaagtcacaagaagtacagcatcgccctggacatcggcaccaactctgtggcgtggccgtg atcaccgacgagtacaaggtgccagcaagaattcaaggtgctgggcaacaccgacccggcacagcatcaa gaagaacctgatcggagccctgctgttcgacagcggcgaacagccgagggccaccggctgaagagaacc gccagaagaagatacaccagacggaagaaccggatctgctatctgcaagagatcttcagcaacgagatggc caaggtggacgacagcttctccacagactggaagagtccttctggtggaagaggataagaagcacgagc ggacccccatcttcggcaacatcgtggacgaggtggcctaccacgagaagtacccaccatctaccacctga gaaagaaactggtggacagcaccgacaaggccgacctgcggctgatctatctggccctggcccatgatc aagttccggggccacttctgatcggaggcgacctgaaccccgacaacagcgacgtggacaagctgttcac cagctggtgcagacctacaaccagctgttcgaggaacccccatcaacgccagcggcggtggacgccaagg ccctcctgtctgccagactgagcaagagcagacggctggaaaatctgatcgccagctgcccggcgagaag aagaatggcctgttcggaaacctgattgccctgagcctgggcctgaccccaactcaagagcaacttcgacc tggccgaggatgccaaactgcagctgagcaaggacacctacgacgacgacgtggacaacctgctggccca gatcgggcaccagtagccgacctgtttctggccgccaagaacctgtccgacgccatcctgctgagcgacat cctgagagtgaacaccgagatcaccaaggccccctgagcgcctctatgatcaagagatacagcagcacc accaggacctgacctgtgaagctctcgtgcggcagcagctgcctgagaagtacaaagagattttctcga ccagagcaagaacggctacccgggtacattgacggcgagccagccaggaagagtctacaagttcatca agcccatcctggaaaagatggacggcaccgaggaactgctcgtgaagctgaacagagaggacctgctgcg gaagcagcggaccttcgacaacggcagcatccccaccagatccacctgggagagctgcacgccattctgc ggcggaaggagattttaccattctgaaggacaacgggaaaagatcgagaagatctgacctccgcat cccctactacgtggggccctctggccaggggaaacagcagattcgctggatgaccagaaagagcgaggaa accatcacccctggaacttcgaggaagtgttggaacaagggcgcttcgcccagagcttatcgagcggatg accaacttcgataagaacctgccaacagagaaggtgctgccaagcacagcctgctgtacgagtactcacc gtgtataacgagctgaccaaagtgaatactgacggagggaatgagaaagccgccttctgagcggcga gcagaaaaaggccatcgtggacctgctgtcaagaccaaccggaaagtaccgtgaagcagctgaaagag gactactcaagaaaatcagtgcttcgactccgtggaatctccggcgtggaagatcggttcaacgcctcct gggcacataccacgatctgctgaaaattatcaaggacaaggacttctggacaataggaaaacgaggacat tctggaagatatcgtgctgacctgacactgtttgaggacagagagatgatcagggaacggctgaaaacctat gccacctgttcgacgacaaaagtgatgaagcagctgaagcggcgagatacaccggctggggcaggctga gccgggaagctgatcaacggcatccgggacaagcagctccggcaagacaatcctggatttctgaagtccgac ggcttcgccaacagaaacttcagcagctgatccacgacgacagcctgaccttaagaggacatccagaaa gccaggtgtccggccaggcgatagcctgcacgagcacattgccaatctggccggcagccccgccattaa gaaggcgatcctgcagacagtgaaggtggtggacgagctcgtgaaagtgtggccggcacaagcccgag aacatcgtgatcgaatggccagagagaaccagaccaccagaaggagacagaagaacagccgcgagaga atgaagcggatcgaagaggcatcaaaagctgggcagccagatcctgaaagaacacccctgtgaaaaca ccagctgcagaacgagaagctgtacctgtactacctgcagaatggcgggatgtacgtggaccagggaac tggacatcaaccggctgtccgactacgatgtggacgctatcgtgcctcagagcttctgaaggacgactccatc gacaacaagtgctgaccagaagcgacaagaaccggggcaagagcgacaacgtgcctccgaagaggtc gtgaagaagtgaagaactactggcgagctgctgaacgccaagctgattaccagagaaagttcgacaat ctgaccaaggccgagagaggcgccctgagcgaactggataaggccggttcacagagacagctggtgg aaaccggcagatcacaagcacgtggcacagatcctgactccggatgaactaagtagcagcagaat gacaagctgatccgggaagtgaagtgatcacctgaagtccaagctggtgtccgatttccgaaggatttcc agttttacaagtgcgcgagatcaacaactaccaccacgcccacgacgcctacctgaacgcccgtcgtgggaa ccgcccctgatcaaaaagtaccctaagctggaaaagcaggttcgtgtacggcgactacaaggtgtacgacgtgc	147

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	ggaagatgatcgccaagagcgagcaggaaatcggaaggctaccgccaagtacttcttacagcaacatca tgaacttttcaagaccgagattaccctggccaacggcgagatccggaagcggcctctgatcgagacaaacg gcgaaacggggagatcgtgtgggataagggccgggattttgccaccgtgcggaagtctgagcatgccc caagtgaatatcgtgaaaaagaccgaggtgcagacagggcggttcagcaagagtcctatcctgccaaag gaacagcgataagctgatcgccagaaaagaaggactgggaccctaagaagtacggcggttcgacagcccc accgtggcctattctgtgctggtggtggccaaagtggaaaaggcaagtccaagaaactgaagagtgtaaa gagctgctggggatcaccatcatggaagaagcagcttcgagaagaatcccatgactttctggaagccaag ggctacaaagaagtgaaaaaggacctgatcatcaagctgcctaagtactccctgttcgagctggaaaacggc cggaagagaatgctggcctctgccggcgaactgcagaagggaacgaactggcctgccctccaatatgt gaacttctgtacctggccagccactatgagaagctgaagggtcccccgaggataatgagcagaaacagct gtttgtggaacagcacaagcactacctggacgagatcatcgagcagatcagcgagttctccaagagagtgatc ctggccgacgctaattcggacaaagtgtgtccgctacaacaagcaccgggataagcccatcagagagca ggcggagaatcatccacctgtttaccctgaccaatctgggagcccctgccgctcaagtactttgacacca ccatcgaccggaagaggtacaccagcaccaaagggtgctggacgccaccctgaccaccagagcatcac cggcctgtacgagacacggatcgacctgtctcagctgggaggtgacgggtcaaaaagaaccccgacggc agcgaattcgagtcaccaagaagaaggaaagtctaa	
SplintR-LZ	atggcgataacgaagccctgttggcagctacgttggaaaatattgaggacgtacagttcccatgccttgccact ccgaagatcgatggaatccgatccgtgaacagacacaaatgcttagcagaacattcaaacccatcaggaac agcgtaatgaatagattgcttacgggaactcttccccgaagggtctgacggcgagattagcatcgaaggagcga ctttccaagatactacctcagcagttatgacgggacacaagatgtataatgctaaattctcatattactggttgact atgttactgacgatccttgaagaaatcatagacagggttgaagatatgaaaaattacataactgtccaccctca tatcctggagcatgcacaggtaaagattatccccgtcataccagtagaaattaacaatataaacgaattgttgca gtatgaacgcgatgtgctctctaaaggcttcgagggcgatgataaggaagcctgatggcaaatataagttcg gtaggtccacattgaaagagggaattctctgaagatgaaacagtttaaggatgcggaagctactatcattagta tgacggcactgtttaaaaacactaacactaaaaccaaggacaactttgctatagtaaaaggagcacacacaa atcaggaaaagttgaggaggacgtaatgggcagtataaggtagattacgatggtgtggtgtttagcattggaa cgggcttcgacgctgaccagcggagggactttggcagaataaggaaagttacattggcaagatggttaatt caaatacttgagatgggctcaaaagactgtccgagatttctgtgtttattggaatcagacacgaagaggatag gAGTGgaggctcaaacggacagcagactcccaacattcaacacccccaaaaacaaagcgaaggttag agtttgagccaaaaaagaaaagaaaggtcGGAGGAGGAGGCAGTGGTGGGCGACT TGAAATTAGAGCCGCGTTCTTGCGCCAGAGGAATACGGCTCTCCG CACGGAGGTAGCCGAACCTTGAGCAAGAAGTACAGAGATTGGAGA ACGAGGTTTCACAGTATGAGACACGATATGGCCCCCTTGCGCGCG GAAAGtaa	148
hLIG4(1-620)-LZ	ATGGCAGCTTCTCAGACCTCTCAAACAGTAGCCTCTCATGTACCGT TCGCTGACTTGTGTTCTACGCTCGAACGCATCCAGAAATCAAAGG GGCGCGCCGAGAAAATCCGGCACTTCAGAGAATTCTTGGATTCTT GGAGGAAGTTTCATGATGCTCTCCACAAAAATCACAAAGATGTAA CGGATAGTTTCTACCCTGCTATGAGACTTATACTGCCGCAGCTTGA GAGGGAACGCATGGCGTATGGTATAAAGGAGACAATGTTGGCGA AATTGTATATTGAGCTGCTGAACTTGCCAAGAGATGGAAAGGACG CGCTCAAACCTGCTGAACTATAGAACACCCACGGGTACCCATGGTG ACGCCGGTGACTTTGCCATGATCGCCTATTTTCGTACTGAAACCTCG ATGTCTTCAAAAAGGTTCTCTTACAATTCAGCAAGTCAACGACCTG CTGGATTCAATTGCGAGTAACAACAGCGCTAAGCGAAAGGATCTC ATTAAGAAAAGCCTCCTGCAGCTGATAACTCAGTCCTCTGCACTC GAACAAAAATGGCTGATTCGGATGATTATCAAGGATTTGAAGTTG GGGGTATCTCAGCAAACCTATTTTCAGCGTGTTTCACAATGATGCAG CAGAATTGCATAATGTCACAACAGATCTTGAGAAAGTCTGCCGAC	149

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	AGTTGCACGACCCCTCTGTAGGCTTGAGTGACATATCTATAACACT TTTTCTGCGTTCAAACCCATGTTGGCTGCTATTGCGGACATAGAA CACATCGAGAAAGACATGAAACATCAGTCATTCTATATAGAGACT AAATTGGACGGCGAGAGGATGCAAATGCACAAAGATGGTGATGT GTATAAATATTTTTCCCGCAACGGCTACAACACTACACTGATCAATTC GGAGCGTCCCCAACTGAAGGGTCCCTCACTCCTTCATACACAAT GCGTTTAAGGCCGATATTCAGATATGTATCCTCGACGGCGAAATG ATGGCGTACAATCCCAATACCCAGACCTTCATGCAAAAGGGAACG AAGTTCGATATTAACCGGATGGTTGAAGATTCCGACCTCCAAACA TGTTACTGTGTGTTTGATGTCCTGATGGTGAATAACAAAAAACTCG GCCATGAAACCCTTCGAAAGCGATACGAAATACTCAGCAGTATAT TACTCCAATACCAGGCCGAATCGAGATCGTACAGAAAACACAAG CCCATACTAAGAATGAAGTTATTGATGCACTGAACGAAGCCATAG ACAAGAGGGAAGAAGGCATAATGGTCAAGCAGCCTCTGAGTATA TATAAACCTGACAAAAGGGGGGAAGGATGGCTGAAGATAAAGCC AGAATACGTGTCTGGTCTTATGGACGAATTGGACATTCTCATCGTC GGAGGATATTGGGGTAAGGGTTCCAGGGGGGGGATGATGTCCAC TTTCTGTGTGCGGTTGCCGAGAAACCGCCCCCAGGGGAAAAACCA TCAGTGTTCCATACGTTGTACGCGTCGGCTCAGGTTGTACGATGA AGGAACTTTACGATCTGGGGTTGAAACTCGCCAAATATTGGAAGC CATTCCATCGGAAAGCACCGCCCTCTAGTATCTTGTGTGGGACGG AGAAGCCAGAAGTTTATATAGAGCCATGTAACCTCAGTAATTGTTC AAATCAAAGCCGCAGAGATCGTCCCGTCAGACATGTACAAGACTG GATGCACCCTTAGATTTCTCGCATCGAAAAATAAGAGATGATA AAGAGTGGCATGAGTGCATGACTCTTGACGACCTTGAACAGCTCC GCGGGAAGGCCAGCGGTAAACTGGCTAGTAAGCACCTCTACATCG GGGGTGACAGTGgaggctccaaacggacagcagactcccaacattcaacccccaaaaa gcggaaggtagagtttgagccaaaaagaaaagaaaggtcGGAGGAGGAGGCAGTGGTG GGCGACTTGAAATTAGAGCCGCGTTCTGCGCCAGAGGAATACGG CTCTCCGCACGGAGGTAGCCGAACTTGAGCAAGAAGTACAGAGAT TGGAGAACGAGGTTTACAGTATGAGACACGATATGGCCCCCTTG GCGGCGGAAAGtaa	
nCas9- hLIG4(1- 620)	atgaaacggacagccgacggaagcgagttcgagtcaccaaaagaagaagcggaaagtcgacaagaagtac agcatcgccctggacatcgccaccaactctgtgggctgggcccgtgacaccgacgagtacaaggtgccag caagaaattcaaggtgctgggcaacaccgaccggcagcagcatcaagaagaacctgatcgagccctgctgtt cgacagcggcgaaacagccgagggcaccggctgaagagaaccgccaagaagaagatacaccagcggga agaaccggatctgctatctgcaagagatctcagcaacgagatggccaaggtggacgacgcttctccacag actggaagagtccttctggtggaagaggataagaagcagcagcggcaccatcttcggcaacatcgtgga cgaggtggcctaccacgagaagtacccaccatctaccacctgagaaagaaactggtggacagcaccgaca aggccgacctgaggctgatctatctggccctggcccatgatcaagttccggggccacttctgatcgaggg cgacctgaaccccgacaacagcagctggacaagctgttcacagctggtgcagacctacaaccagctgtt cgaggaaaaccccatcaacgcccagcggcgtggacccaagccatcctgtctgccagactgagcaagagc agacggctggaaaatctgatcgcccagctgcccggcgagaagaagaatggcctgttcggaacctgattgcc ctgagcctgggctgaccccaactcaagagcaacttcgacctggccgaggatgccaactgcagctgagc aaggacacctacgacgacgacctggacaacctgctggcccagatcgccgaccagtacgccacctgtttctg gccgcaagaacctgtccgacgcatctctgagcgacatctgagagtgaacaccgagatccaagggc ccccctgagcgctctatgatcaagagatacagcagcaccaccagcagctgacctgctgaaagctctcgt gcggcagcagctgctgagaagtacaaagagattttctcgaccagagcaagaacggctacgcccgtacat tgacggcgagccagccaggaagagttctacaagttcatcaagccatcctggaaaagatggacggcaccg	150

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	aggaactgctcgtgaagctgaacagagaggacctgctgcggaagcagcggaccttcgacaacggcagcat cccccaccagatccacctgggagagctgcacgccattctgcggcggcagggaagattttaccattcctgaag gacaaccgggaaaagatcgagaagatcctgaccttcgcacccctactacgtgggacctctggccagggg aaacagcagattcgcttgatgaccagaaaagcagcaggaaaccatcacccctggaaacttcgaggaaagg tggacaaggcgcttcgccagagcttcacgagcggatgaccaacttcgataagaacctgcccacagaga aggtgctgcccaagcacagcctgctgtacgagtacttcaccgtgtataacgagctgaccaaagtgaatacgt gaccgagggaaatgagaagccccgcttcctgagcggcagcagaaaaaggccatcgtggacctgctgttca agaccaaccggaaagtgaccgtgaagcagctgaagaggactacttcaagaaaatcgagtgttcgactccg tggaaatctcggcgtggaagatcggttcaacgcctccctgggcacataccacgatctgctgaaaattatcaag gacaaggacttctggacaatgaggaacgaggacattctggaagatatcgtgctgacctgacactgtttg aggacagagagatgatcaggaacggctgaaaacctatgccacctgttcgacgacaaaagtgatgaagcag ctgaagcggcggagataccggcgtggggcaggctgagccggaagctgatcaacggcatccgggacaag cagtccggcaagacaatcctggatttctgaagtccgacggcttcgccacagaaaacttcagctgatcc acgacgacagcctgacctttaagaggacatccagaaagcccagggttcggccaggggcgatagcctgcac gagcacttggcaatcggccggcagccccgccattaagaaggccatcctgcagacagtgaaggtggtgga cgagctcgtgaaagtgtggccggcacaagcccagagaacatcgtgatgaaatggccagagagaaccag accaccagaaagggacagaagaacagccgagagagaatgaagcggatcgaagaggccatcaagagct gggcagccagatcctgaaagaacccccgtggaacacccagctgcagaacgagaagctgtacctgtact acctgcagaatggcggggatgtacgtggaccaggaactggacatcaaccggctgtccgactacgatgtgg acgtatcgtgctcagagcttctgaaggacgactccatcgacaacaaggtgctgaccagaagcgacaaga accggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaactactggcgcgagct gctgaacgccaagctgattaccagagaaagtgcacaatctgaccaaggccgagagaggcgccgctgagc gaactggataaaggccggttcacaaagacagctggttgaaacccggcagatcacaaagcagctggcaca gatcctggactcccggatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaagtgatcac cctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaagtgcgcgagatcaacaactacc accacgcccacgacgctacctaagcgcgtcgtgggaaccgccctgatcaaaaagtaccctaagctggaa agcaggttcgtgtacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcagcaggaaat cggcaaggctaccgccaagtacttctctacagcaacatcatgaacttttcaagaccgagattacctggccaa cggcgagatccggaagcggcctctgatcgagacaaacggcgaaacccggggagatcgtgtgggataaggg ccgggattttgccaccgtcgggaagtgtgagcatgccccaaagtgaatatcgtgaaaaagaccgaggtgca gacaggcggcttcagaaaggtctatcctgcccaagaggaacagcgataagctgatcgccagaaagaagg actgggaccctaagaagtacggcggcttcgacagccccaccgtggcctattctgtgctggtggtggccaaagt ggaaaaggcgaagtccaagaactgaagagtgtaagagctgctggggatcaccatcatggaagaagc agcttcgagaagaatccatcgacttctggaagccaagggtacaaagaagtgaagaaggacctgatcatca agctgcctaagtactccctgttcgagctggaaaacggccgggaagagaatctggcctctgcccggcgaactgc agaagggaacgaactggcctgcctcccaatatgtgaacttctgtacctggccagccactatgagaagct gaagggtcccccgaggataatgagcagaaacagctgtttgtggaacgcacaagcactacctggacgaga tcatcgagcagatcagcaggttctcaagagagtgatcctggccgacgctaacttgacaaaagtgtgtccgc ctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccacctgtttacctgaccaat ctgggagcccctgcccctcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccaaaga ggtgtgagcgcaccctgatccaccagagcatcaccggcctgtacgagacagggatcgacctgtctcagct gggaggtgacTCCGGTGGCTCCTCAGGGGGATCTAAACGCACGGCCGA TGGGTCCGAGTTTGAGTCTCCCAAGAAGAAAAGGAAAGTGAGTGG TGGAAGTAGCGGCGGTAGCGCAGCTTCTCAGACCTCTCAAACAGT AGCCTCTCATGTACCGTTCGCTGACTTGTGTTCTACGCTCGAACGC ATCCAGAAATCAAAGGGGCGCGCCGAGAAAATCCGGCACTTCAG AGAATTCTTGGATTCTTGGAGGAAGTTTCATGATGCTCTCCACAAA AATCACAAAGATGTAACGGATAGTTTCTACCCTGCTATGAGACTT ATACTGCCGCAGCTTGAGAGGGAACGCATGGCGTATGGTATAAAG	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	GAGACAATGTTGGCGAAATTGTATATTGAGCTGCTGAACTTGCCA AGAGATGGAAAGGACGCGCTCAAACCTGCTGAACTATAGAACACC CACGGGTACCCATGGTGACGCCGGTGACTTTGCCATGATCGCCTA TTTCGTA CTGAAACCTCGATGTCTTCAAAAAGGTTCTCTTACAATT CAGCAAGTCAACGACCTGCTGGATTCAATTGCGAGTAACAACAGC GCTAAGCGAAAGGATCTCATTAAAGAAAAGCCTCCTGCAGCTGATA ACTCAGTCCTCTGCACTCGAACAAAAATGGCTGATTTCGGATGATT ATCAAGGATTTGAAGTTGGGGGTATCTCAGCAAACTATTTTCAGC GTGTTTCACAATGATGCAGCAGAATTGCATAATGTCACAACAGAT CTTGAGAAAGTCTGCCGACAGTTGCACGACCCCTCTGTAGGCTTG AGTGACATATCTATAACACTTTTTTCTGCGTTCAAACCCATGTTGG CTGCTATTGCGGACATAGAACACATCGAGAAAGACATGAAACATC AGTCATTCTATATAGAGACTAAATTGGACGGCGAGAGGATGCAAA TGCACAAAGATGGTGATGTGTATAAATATTTTCCCGCAACGGCT ACAACTACACTGATCAATTCGGAGCGTCCCCAACTGAAGGGTCCC TCACTCCTTTCATACACAATGCGTTTAAAGGCCGATATTCAGATATG TATCCTCGACGGCGAAATGATGGCGTACAATCCCAATACCCAGAC CTTCATGCAAAAGGGAACGAAGTTCGATATTAAACGGATGGTTGA AGATTCCGACCTCCAAACATGTTACTGTGTGTTTGATGTCCTGATG GTGAATAACAAAAAACTCGGCCATGAAACCCCTTCGAAAGCGATAC GAAATACTCAGCAGTATATTTACTCCAATACCAGGCCGAATCGAG ATCGTACAGAAAACACAAGCCATACTAAGAATGAAGTTATTGAT GCACTGAACGAAGCCATAGACAAGAGGGAAGAAGGCATAATGGT CAAGCAGCCTCTGAGTATATATAAACCTGACAAAAGGGGGGAAG GATGGCTGAAGATAAAGCCAGAATACGTGTCTGGTCTTATGGACG AATTGGACATTCTCATCGTCGGAGGATATTGGGGTAAGGGTTCCA GGGGGGGGATGATGTCCCACCTTTCTGTGTGCGGTTGCCGAGAAAC CGCCCCCAGGGGAAAAACCATCAGTGTTCCATACGTTGTACGCG TCGGCTCAGGTTGTACGATGAAGGAACCTTACGATCTGGGGTTGA AACTCGCCAAATATTGGAAGCCATTCCATCGGAAAGCACCGCCCT CTAGTATCTTGTGTGGGACGGAGAAGCCAGAAGTTTATATAGAGC CATGTAACCTCAGTAATTGTTCAAATCAAAGCCGCAGAGATCGTCC CGTCAGACATGTACAAGACTGGATGCACCCTTAGATTTCTCTCGCAT CGAAAAAATAAGAGATGATAAAGAGTGGCATGAGTGCATGACTC TTGACGACCTTGAACAGCTCCGCGGGAAGGCCAGCGGTAAACTGG CTAGTAAGCACCTCTACATCGGGGGTGACtaa	
T4-nCas9	atgatgacctaagattctcaacgaaatcgtagtatagggtccactaagcagaagcaggccatattggaaaa aaataaggacaatgaactttgaagagagtctatagactgacgtactctagggggctccagtactacatcaaga aatggccta aacctggcattgacgacgcagtcattcggtatgctgacattgacggatatgttgatttcattgagttt acgctggccaccagaaaacttacgggtaattgctgcatagaagaacttacagggtacataacagacgggaa gaaagatgacgtggaagtgtcagacgagttatgatgcgcatctcagtgaggcgctagcgtgtcaatcgc gaacaaagtctggcccggcctcataccagagcagccacagatgctggcatcttctatgacgaaaaaggcat aaacaagaatattaaattcccggccttcgctcaactcaaagcagatggtgccagggtgtttgccgaagtgcggg gtgatgaacttgatgacgtgcggctctgtctagggcaggtaacgagtacctcgccctggacttgcttaagag gaactgattaaatgacagctgaggcgcgagataaccccgaggcgctctatcgacggggagctggt gtatcacgaacaagttaaaaggaaccggagggtcttgattttctttcagcgctatcctgagaacagcaagg cgaaagaatttcagaaagtcagaaagcaggaccgcaagtaaatggaatcgtaataaaagcctcaagggtta ccatcagcgaaaaagaagccagtgcatgaaattcaagttgggactatgtcccttggtcgaaatttactccc tgcccgcatccggctgaagtatgatgttcgcttcagtaaaactggagcaaatgacgagcggttatgataaggtta	151

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	tacttattgagaatcaggtcgtaaataatttggacgagggcgaaaagtatatatacaaaaagtatatagaccaagggtt ggaggggatcattttgaagaacatagacggactttgggagaacgcccgggtccaagaattgtataaattcaaag aagtcatagatgttgacctcaagatagtaggtatataatccccacagaaaggaccaaccaaaagcaggcggatt cattttggagtcaggagtggtgggaagataaaggtaaatgctggatctggactcaaggacaagctggtgtgaag tcacatgaactggaccgaaccaggattatggagaatcagaactattacatcgggaagatattggagtgtaag caacggctggcttaaatcagatggagaactgattacgttaaatgttctgcccatagccatacactccgcga ggacaaaacgaaggctaacacgtttgaagacgtattcggagatttccatgaggtgactggcctttccggtggct cctcagggggatctaaacgcacggccgatgggtccgagtttgagctctccaagaagaaaaggaaagtgag ggtggaagtagcggcggtagcgacaagaagtacagcatcggtcggacatcggcaccaactctgtgggctg ggcgtgatcaccgacgagtacaaggtgcccagcaagaattcaaggtgctgggcaacaccgaccggcac agcatcaagaagaacctgatcggagccctgctgttcgacagcggcgaaacagccgagggccaccggctga agagaaccgccaagaagaagataccagacggaagaaccggatctgctatctgcaagagatcttcagcaac gagatggccaagggtggacgacagcttctccacagactggaagagtccttctggtggaagaggataagaag cacgagcggcaccatcttcggcaacatcgtggacgaggtggcctaccacgagaagtacccccaccatcta ccactgagaaaagaaactggtggacagcaccgacaaggccgacctgcggtgatctatctggccctggccc acatgatcaagttccggggccacttctgatcagggcgacctgaaccccgacaacagcgacgtggacaag ctgttcacccagctggtgacagactacaaccagctgttcgaggaacccccatcaacgccagcgcggtggac gccaaggccatctgtctgacagactgagcaagagcagacggctggaaaatctgatcgcccagctggccgg cgagaagaagaatggcctgttcggaacactgattgccctgagcctgggctgaccccaacttcaagagcaa cttcgacctggccgaggatgccaactgcagctgagcaaggacacctacgacgacgacctggacaacctgc tggcccagatcgcgaccagtacgcccactgttctggccgccaagaacctgtccgacgccatcctgctga gcgacatcctgagagtgaacaccgagatcacaaggccccctgagcgctctatgatcaagagatacagac gagcaccaccaggacctgacctgctgaaagctctcgtgcggcagcagctgcctgagaagtacaagagat tttcttcgaccagagcaagaacggctacgcccgtacattgacggcgagccagccaggaagagtctacaa gttcatcaagcccactcctggaaaagatggacggcaccgaggaactgctcgtgaagctgaacagagaggacc tgctgcggaagcagcgaccttcgacaacggcagcatccccaccagatccactgggagagctgcacgc cattctgcggcgaggaagattttaccattcctgaaggacaaccgggaaaagatcgagaagatcctgacc ttccgcatcccctactacgtggccctctggccaggggaaacagcagattcgctggatgaccagaagagc gaggaaacatcacccccctggaacttcgaggaagtgggtggacaaggcgcttcgcccagagcttcatcga gcggtgatcaacttcgataagaactcgccaacgagaaggtgctgccaagcacagcctgctgtacgagta cttcacgtgtataacgagctgacaaagtgaatacgtgaccgaggggaatgagaaagccgcttctgag cggcgagcagaaaaagccatcgtggacctgctgtcaagaccaaccggaaagtgacctgaagcagctg aaagaggactactcaagaaatcagtgcttcgactcctggaaatctccggcgtggaagatcggttcaacg cctccctgggacataccacgatctgctgaaattatcaaggacaaggacttcttgacaatgaggaaaacga ggacattctggaagatatcgtgctgacctgacactgtttgaggacagagagatgatcaggaacggctgaaa acctatgccacctgttcgacgacaagtgtgaagcagctgaagcggcgagatacaccgctggggcag gctgagccggaagctgatcaacggcatccgggacaagcagtcgggcaagacaatctggatttctgaagtc cgacggcttcgccaacagaaacttcatgcagctgatccacgacgacgacctgacctttaaaggacatcca gaaagcccaggtgtccggccaggcgatagcctgcacgagcacattgccaatctggccggcagccccgcc attaagaaggcatcctgcagacagtgaaggtggtggacgagctcgtgaaagtgtggccggcacaagcc cgagaacatcgtgatgaaatggccagagagaaccagaccaccagaaggagcagaagaacagccgca gagaatgaagcggatcgaagaggcatcaagagctgggagccagatcctgaaagaacacccgtggaa aacacccagctgcagaacgagaagctgtacctgtactacctgcagaatggcgggatgtacgtggaccag gaactggacatcaaccgctgtccgactacgatgtggacgctatcgtgcctcagagcttctgaaggacgact ccatcgacaacaaggctgtgaccagaagcgacaagaaccggggcaagagcgacaacgtgcctccgaag aggtcgtgaagaagatgaagaactactggcgagctgctgaacccaagctgattaccagagaaagtgcg acaatctgaccaaggccgagagaggcggtgagcgaactggataaggccggttcatcaagagacagct gggtgaaaccggcagatcacaagcacgtggcacagatcctggactccggatgaactaagtacgacg agaatgacaagctgatccgggaagtgaagtgtatcacctgaagtccaagctggtgtccgatttccggaagg	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	atttcagttttacaaagtgcgcgagatcaacaactaccaccacgcccacgacgcctacctgaacgccgtcgt gggaaccgccctgatcaaaaagtaccctaagctggaaagcgagttcgtgtacggcgactacaaggtgtacga cgtcgcgaagatgatcgccaagagcgagcaggaaatcggaaggctaccgccaagtacttcttacagca acatcatgaactttttcaagaccgagattaccctggccaacggcgagatccggaagcggcctctgatcgagac aaacggcgaaaccggggagatcgtgtgggataaggcggggattttgccaccgtgcggaagtgtgagc atgccccaaagtgaatatcgtgaaaaagaccgaggtgcagacaggcggttcagcaaaagagtctatcctgcc aagaggaacagcgataagctgatcgccagaaagaaggactgggaccctaagaagtacggcggttcgaca gccccaccgtggcctattctgtgtgtgtggccaaagtggaaaagggaagtcgaagaactgaagagtgtg tgaagagctgctggggatcacatcatggaagaagcagcttcgagaagaatcccatgacttttggaaagc caagggtacaaagaagtgaaaaaggacatgatcatcaagctgcctaagtactcctgttcgagctggaaac ggccgggaagagaatgctggcctctgccggcgaaactgcagaagggaacgaactggccttgccctcaa atatgtgaacttctgtacctggccagccactatgagaagctgaagggtccccggaggataatgacagaaac agctgtttgtggaacagcacaagcactacctggacgagatcatcgagcagatcagcgagtttccaagagagt gatcctggccgacgctaacttgacaaaagtgtgtccgctacaacaagcaccgggataagcccatcagaga gcaggccgagataatcatccactgtttaccctgaccaatctgggagccccctggccttcaagtactttgaca ccaccatcgaccggaagaggtacaccagcacaagaggtgctggacgccaccctgatccaccagagcat caccggcctgtacgagacacggatcgacctgtctcagctgggaggtgacaaacggacagccgacggaagc gagttcgagtcacaaagaagaagcggaagtctaa	
SplintR-nCas9	atgatggcgataacgaagccctgtgtggcagctacgttggaatattgaggacgtacagttcccatgccttgc cactccgaagatcgatggaatccgatccgtgaacagacacaaatgcttagcagaacattcaaaccatcagg aacagcgtaataatgattgttaccggaactcttggccgaagggtctgacggcgagattgacatcggaaggag cgactttccaagatactacctcagcagttatgacgggacacaaagatgtataatgctaaattctcatattactgttt gactatgttactgacgatcctttgaagaatacatagacagggttgagatatgaaaaattacataactgtccacc ctcatatcctggagcatgcacaggtaaagattatcccgctcataccagtagaaattaacaatataaccgaattgtt gcagtatgaacgcgatgtgtctctaaaggcttcgagggcgtgatgataagggaagcctgatggcaatataag ttcggtaggtccacattgaaagagggaattcttgaagatgaaacagtttaaggatgcggaagctactatcatt agtatgacggcactgtttaaaaacactaaactaaaaccaaggacaactttggctatagtaaaaggagcacac acaaatcaggaaaagttgaggaggacgtaatgggcagtatagaggtagattacgatggtgtgtgtttagcatt ggaacgggcttcgacgtgaccagcggaggacttttggaagaataaggaaagtacattggcgaagatggtt aaattcaataacttcgagatgggctcaaaagactgtccgagatttctgtgtttattggaatcagacacgaagag gataggtCCGGTGGCTCCTCAgggggatctaaacgcacggccgatgggtccgagtttgatctc ccaagaagaaaaggaaagtgagtggtggaagtgcggcgtagcgacaagaagtacagcatcggcctgga catcgccaccaactctgtgggctgggcccgtgatcaccgacgagtagaaggtcccagcaagaattcaaggt gctgggcaacaccgaccggcacagcatcaagaagaacctgatcgagccctgctgttcgacagcggcgaa acagccgaggccaccggctgaagagaaccgccaagaagatacaccagacggaagaaccggatctgc tatctgcaagagatctcagcaacgagatggccaaggtggacgacagcttctccacagactggaagagtcct tcctggtggaagaggataagaagcacgagcggcaccatcttcggcaacatcgtggacgaggtggcctac cacgagaagtacccaccatctaccacctgagaagaactgtgtggacagcaccgacaaggccgacctgc ggctgatctatctggcctggccacatgatcaagttccggggccacttctgatcaggggcagctgaaccc cgacaacagcgacgtggacaagctgttcacagctggtgcagacctacaaccagctgttcgaggaacccc catcaacgccagcggcgtggacgccaaggccatcctgtctgccagactgagcaagagcagacggctggaa aatctgatcgccagctgcccggcgagaagaagaatggcctgttcgaaacctgattgccctgagcctgggc ctgaccccaactcaagagcaacttcgacctggccgaggatgcaaaactgcagctgagcaaggacacctac gacgacgacctggacaacctgctggccagatcggcgaccagtacggcagctgttctggccgcaagaa cctgtccgacgccatcctgtgagcgacatcctgagagtgaaaccgagatcaccaaggccccctgagcg cctctatgatcaagagatacagcagcaccaccagacctgacctgctgaaagctctcgtcggcgacgagc tgcctgagaagtacaaagagatttctcgaccagagcaagaacggctacggcgctacattgacggcgag ccagccggaagagttctacaagttcatcaagccatcctggaaaagatggacggcaccgaggaactgctcg tgaagctgaacagagaggacctgctgcggaagcagcgaccttcgacaacggcagcatccccaccagat	152

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	ccacctgggagagctgcacgccattctgcggcggcaggaagattttaccattctgaaggacaaccggga aaagatcgagaagatcctgacctccgcatcccctactacgtgggcccctctggccaggggaacagcagatt cgcttgatgaccagaaaagcgcaggaaccatcacccctggaacttcgaggaagtggaggacaagggc gctccgcccagagcttcacgagcggatgaccaacttcgataagaacctgcccacgagaaggtgctgcc aagcacagcctgctgtacgagtacttcacgtgtataacgagctgaccaaagtgaatacgtgaccgagggg atgagaaaagcccgccttctgagcggcgcagagaaaaggccatctggacctgctgttcaagaccaaccg gaaagtgaccgtgaagcagctgaagaggactacttcaagaaaatcagtgcttcgactccgtggaatctcc ggcgtggaagatcggttaacgcctccctgggcacataccacgatctgctgaaaattatcaaggacaaggact tctggacaatgaggaacagaggacattctggaagatatcgtgctgacctgacactgtttgaggacagaga gatgatcaggaacggctgaaaacctatcccacctgttcgacgacaagtgatgaagcagctgaagcggc ggagatacaccggctggggcaggtgagccggaagctgatcaacggcatccgggacaagcagctccggca agacaatcctggatttctgaagtccgacggcttcgccaacagaaacttcagcagctgatccacgacgacag cctgacctttaagaggacatccagaaagcccaggtgtccggccagggcgatagcctgcacgagcacattg ccaatctggcggcagccccgccattaagaagggcacatctgcagacagtgaaggtggtggacgagctcgtg aaagtgtggcggcgcacaagcccagaaacatcgtgatcgaatggccagagagaaccagaccaccaga agggacagaagaacagccgagagagaatgaagcggatcgaagaggccatcaagagctgggcagccag atcctgaagaaacaccccggtgaaaaacacccagctgcagaacgagaagctgtacctgtactacgtcagaat gggcccggatgtacgtggaccaggaactggacatcaaccggctgtccgactacgatgtggacgctatcgtg cctcagagcttttgaaggacgactccatcgacaacaagggtgctgaccagaagcgacaagaacccgggcaa gagcgacaacgtgcctccgaagaggtcgtgaagaagatgaagaactactggcggcagctgctgaacgcc aagctgattacccagagaaagttcgacaatctgaccaaggccgagagaggcggcctgagcgaactggataa ggccggcttcataagagacagctggtggaacccggcagatcacaaagcacgtggcacagatcctggact cccggtatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaagtgatcacctgaagtcca agctggtgtccgatttccggaaggatttccagttttacaagtgcgcgagatcaacaactaccaccacgcccac gacgcctacctgaacgccgtcgtgggaaccgcccctgatcaaaaagtaccctaagtggaaagcgagttcgtg tacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcgcagcaggaatcggaaggcta ccgccaagtaacttcttacagcaacatcatgaactttttcaagaccgagattaccctggccaacggcgagatcc ggaagcggcctctgatcgagacaacggcgaaaccggggagatcgtgtgggataaggggccgggattttgc caccgtgcggaaagtgtgtagcatgccccagtgaaatcgtgaaaaagaccgaggtgcagacagggcggt tcagcaaaagagtctatcctgccaaagaggaaacagcgataagctgatcgccagaaagaaaggactgggacct aagaagtacggcggttcgacagccccaccgtggcctattctgtgctggtggtggccaaagtggaaaagggc aagtccaagaaactgaagagtgtgaagagctgctggggatcaccatcatggaagaaagcagcttcgagaa gaatcccatcgacttctggaagccaagggtacaaagaagtgaagaaaggacatcatcaagctgcctaag tactccctgttcgagctggaaaacggccggaagagaatgctggcctctgcggcgaaactgcagaagggaaa cgaactggccctgccctccaatatgtgaacttctgtacctggccagccactatgagaagctgaagggtcc cccgaggataatgagcagaaacagctgtttgtggaacagcacaagcactacctggacgagatcatcgagca gatcagcaggttccaagagagtgtcctggccgacgctaacttgacaaaagtgtgctccgctacaacaag caccgggataagcccatcagagagcaggccgagaatatcatccacctgtttaccctgaccaatctgggagcc cctgccgcttcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccagagaggtgctggac gccacctgatccaccagagcatcaccggcctgtacgagacacggatcgacctgtctcagctgggaggtga caaacggacagccgacggaagcgagttcagtcaccaaagaagaagcggaaggtctaa	
hLIG4(1-620)-nCas9	ATGGCAGCTTCTCAGACCTCTCAAACAGTAGCCTCTCATGTACCGT TCGCTGACTTGTGTTCTACGCTCGAACGCATCCAGAAATCAAAGG GGCGCGCCGAGAAAATCCGGCACTTCAGAGAATTCTTGGATTCTT GGAGGAAGTTTCATGATGCTCTCCACAAAAATCACAAAGATGTAA CGGATAGTTTCTACCCTGCTATGAGACTTATACTGCCGCAGCTTGA GAGGGAACGCATGGCGTATGGTATAAAGGAGACAATGTTGGCGA AATTGTATATTGAGCTGCTGAACTTGCCAAGAGATGGAAAGGACG CGCTCAAACCTGCTGAACTATAGAACACCCACGGGTACCCATGGTG	153

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	ACGCCGGTGACTTTGCCATGATCGCCTATTTTCGTAAGTCAACGACCTG ATGTCTTCAAAAAGGTTCTCTTACAATTCAGCAAGTCAACGACCTG CTGGATTCAATTGCGAGTAACAACAGCGCTAAGCGAAAGGATCTC ATTAAGAAAAGCCTCCTGCAGCTGATAACTCAGTCCTCTGCACTC GAACAAAAATGGCTGATTCGGATGATTATCAAGGATTTGAAGTTG GGGGTATCTCAGCAAACTATTTTCAGCGTGTTTCACAATGATGCAG CAGAATTGCATAATGTCACAACAGATCTTGAGAAAGTCTGCCGAC AGTTGCACGACCCCTCTGTAGGCTTGAGTGACATATCTATAACACT TTTTCTGCGTTCAAACCCATGTTGGCTGCTATTGCGGACATAGAA CACATCGAGAAAGACATGAAACATCAGTCATTCTATATAGAGACT AAATTGGACGGCGAGAGGATGCAAATGCACAAAGATGGTGATGT GTATAAATATTTTCCCGCAACGGCTACAACACTACACTGATCAATTC GGAGCGTCCCCAACTGAAGGGTCCCTCACTCCTTTCATACACAAT GCGTTTAAGGCCGATATTCAGATATGTATCCTCGACGGCGAAATG ATGGCGTACAATCCCAATACCCAGACCTTCATGCAAAAGGGAACG AAGTTCGATATTAACCGGATGGTTGAAGATTCCGACCTCCAAACA TGTTACTGTGTGTTTGATGTCCTGATGGTGAATAACAAAAAACTCG GCCATGAAACCCTTCGAAAGCGATACGAAATACTCAGCAGTATAT TTAATCCAATACCAGGCCGAATCGAGATCGTACAGAAAACACAAG CCCATACTAAGAATGAAGTTATTGATGCACTGAACGAAGCCATAG ACAAGAGGGAAGAAGGCATAATGGTCAAGCAGCCTCTGAGTATA TATAAACCTGACAAAAGGGGGGAAGGATGGCTGAAGATAAAGCC AGAATACGTGTCTGGTCTTATGGACGAATTGGACATTCTCATCGTC GGAGGATATTGGGGTAAGGGTTCCAGGGGGGGGATGATGTCCAC TTTCTGTGTGCGGTTGCCGAGAAACCGCCCCCAGGGGAAAAACCA TCAGTGTTCATACGTTGTACGCGTCGGCTCAGGTTGTACGATGA AGGAACTTTACGATCTGGGGTTGAACTCGCCAAATATTGGAAGC CATTCCATCGGAAAGCACCGCCCTCTAGTATCTTGTGTGGGACGG AGAAGCCAGAAGTTTATATAGAGCCATGTAACCTCAGTAATTGTTC AAATCAAAGCCGCAGAGATCGTCCCGTCAGACATGTACAAGACTG GATGCACCCTTAGATTTCTCCTCGCATCGAAAAATAAGAGATGATA AAGAGTGGCATGAGTGCATGACTCTTGACGACCTTGAACAGCTCC GCGGGAAGGCCAGCGGTAAACTGGCTAGTAAGCACCTCTACATCG GGGGTGACTCCGGTGGCTCCTCAGGGGGATCTAAACGCACGGCCG ATGGGTCCGAGTTTGAGTCTCCCAAGAAGAAAAGGAAAGTGAGTG GTGGAAGTAGCGGCGGTAGCgacaagaagtacagcatcggcctggacatcgccaccaac tctgtgggctgggcccgtatcaccgacgagtacaagggtgccagcaagaaattcaagggtgctgggcaacac cgaccggcacagcatcaagaagaacctgatcggagccctgctgttcgacagcggcgaaacagccgagggc accgggctgaagagaaccgccaagaagaatacaccagacggaagaaccggatctgctatctgcaagaga tcttcagcaacgagatggccaagggtggacgacagcttctccacagactggaagagtccttctggtggaaga ggataagaagcacgagcggcaccatcttcggcaacatcgtggacgaggtggcctaccacgagaagtacc ccaccatctaccacctgagaagaactgggtggacagcaccgacaaggccgacctgcggtgatctatctgg ccctggcccacatgatcaagttccggggccacttctgatcagggcgacctgaaccccgacaacagcgac gtggacaagctgttcacagctggtgcagacctacaaccagctgttcgaggaaaacccatcaacgccagc ggctggacgccaaggccatctgtctgccagactgagcaagagcagacggctggaaaatctgatcgcca gctgcccgcgagaagaagaatggcctgttcgaaacctgattgccctgagcctgggctgaccccaact caagagcaacttcgacctggcggaggtgccaactgcagctgagcaaggacacctacgacgacgacctg gacaacctgctggccagatcgccgaccagtacggcgacctgttctggccgccaagaacctgtccgacgac atcctgctgagcgacatcctgagagtgaacaccgagatcaccaaggccccctgagcgctctatgatcaag	

Name	Fusion protein polypeptide sequence or nucleic acid sequence	SEQ ID NO:
	agatacgacgagcaccaccaggacctgacacctgctgaaagctctcgtcggcagcagctgcctgagaagta caaagagattttctcgaccagagcaagaacggctacgccggctacattgacggcggagccagccaggaa agttctacaagttcatcaagcccacatctggaaaagatggacggcaccgaggaactgctcgtgaagctgaaca gagaggacctgctcgggaagcagcggaccttcgacaacggcagcatccccaccagatccacctgggaga gctgcacgccattctcggcggcaggaagattttaccattctgaaggacaaccgggaaaagatcgagaa gatcctgaccttcgcacccccactacgtgggcccctctggccaggggaaacagcagattcgctggatgacc agaaaagcgcgaggaaaccatcacccctggaacttcgaggaagtggtggacaagggcgcttcgccccaga gcttcacgcagcggatgaccaacttcgataagaacctgcccacgagaagtgctgccaagcacagcctgc tgtacgagtacttcacgtgtataacgagctgacaaaagtgaatacgtgaccgaggggaatgagaaagcccg ccttctgagcggcgcagagaaaaaggccatcgtggacctgctgttaacgacaaccggaaaagtacccgtg aagcagctgaaagaggactacttcaagaaaatcgagtgttcgactccgtggaaatctccggcgtggaagtc ggttcaacgcctccctgggcacataccacgatctgctgaaattatcaaggacaaggacttctggacaatga ggaaaacgaggacattctggaagatacgtgctgaccctgacactgtttgaggacagagatgatcgagga acggctgaaaacctatgccacacgttcgacgacaaaagtgatgaagcagctgaagcggcggagatacaccg gctggggcagcgtgagccgggaagctgatcaacggcatccgggacaagcagctccggcaagacaatctgga ttctctgaagtcggacggcttcgccaacagaaacttcagcagctgatccacgacgacacccgtgacctttaaag aggacatccagaaagcccagggtgtccggccaggcgatagcctgcacgagcacattgccaatctggccgg cagccccgccattaagaagggcacatctgcagacagtgaaagtggtggacgagctcgtgaaagtgtggggc ggcacaagcccagagaacatcgtgatcgaatggccagagagaaccagaccaccagaaggggacagaaga acagccgcgagagaatgaagcggatcgaaagggcatcaagagctggggcagccagatctgaaagaac acccctgaggaaaacacccagctgcagaacgagaagctgtacctgtactacctgcagaatggggcggtatgt acgtggaccaggaactggacatcaaccggctgtccgactacgatgtggacgctatcgtgcctcagagcttct gaaggacgactccatcgacaacaaggtgctgaccagaagcgacaagaaccggggcaagagcgacaacgt gccctccgaagaggtcgtgaagaagatgaagaactactggcggcagctgctgaacgccaaagctgattacc agagaaggttcgacaatctgaccaaggccgagagaggcggcctgagcgaactggataaggccggcttcac aagagacagctggtggaacccggcagatcacaaagcacgtggcacagatcctggactcccgatgaaca ctaagtacgacgagaatgacaagctgatccgggaagtgaagtgatcacctgaagtccaagctggtgtccg atttccggaaggatttccagttttacaaagtgcgcgagatcaacaactaccaccacgcccacgacgcctacctg aacgcctgctgggaaccgccctgatcaaaagtaccctaagctggaaagcgagttcgtgtacggcgactac aaggtgtacgactgctgggaagatgatcgccaagagcgagcaggaaaatcggaaggctaccgccaagtactt ctctacagcaacatcatgaacttttcaagaccgagattacctggccaacggcgagatccggaagcggcct ctgatcgagacaacggcgaaacggggagatcgtgtgggataaggcgccgggattttgccaccgtgcggaa agtgcctgagcatgcccgaagtgaatatcgtgaaaagaccgaggtgcagacagggcgttcagcaaaagat ctatcctgccaagaggaacagcgataagctgatcgccagaagaaggactgggaccctaagaagtacggc ggcttcgacagccccaccgtggcctattctgtgctggtggtggccaaagtggaaaagggaagtccaagaaa ctgaagagtgtaagagctgctggggatcacatcatggaaagaagcagcttcgagaagaatccatcgac tttctggaagccaagggtacaaagaagtgaaaaaggacctgatcatcaagctgcctaagtactccctgtcga gctggaaaacggccggaagagaatcgtggcctctgcggcgcaactgcagaagggaacgaactggccctg ccctccaaatatgtgaacttctgtacctggccagccactatgagaagctgaagggtcccccgaggataatg agcagaaacagctgtttgtggaacagcacaagcactacctggacgagatcatcgagcagatcagcgattct ccaagagagtgatcctggccgacgctaactctggacaaagtgtgtccgctacaacaagcaccgggataagc ccatcagagagcagggcagataatcatccactgtttaccctgaccaatctgggagccccctgccctcaa gtactttgacaccaccatcgaccggaagaggtacaccagcaccaaagaggtgctggacgccaccctgatcc accagagcatcaccggcctgtacgagacacggatcgacctgtctcagctgggaggtgacaaacggacagcc gacggaagcgagttcgagtcaccaaagaagaagcggaaagtctaa	

[00147] Disclosed herein are protein complexes comprising: an RNA-guided endonuclease bound to a ligase. The endonuclease and the ligase may be bound together through

heterodimerization domains. The heterodimerization domains may include one or more of leucine zippers, PDZ domains, streptavidin and streptavidin binding protein, foldon domains, hydrophobic polypeptides, an antibody that binds the Cas nickase, or an antibody that binds the ligase, or one or more binding fragments thereof.

[00148] In some aspects, the system comprises at least one donor strand. In some aspects, the donor strand comprises a nucleic acid sequence that is at least partially homologous to the genomic locus targeted by the at least one guide nucleic acid. In some aspects, the donor strand comprises a nucleic acid sequence that is not homologous to the genomic locus targeted by the at least one guide nucleic acid. In some aspects, the donor strand is a single-stranded or a double-stranded nucleic acid. In some aspects, the donor strand comprising double-stranded nucleic acid comprises at least one overhang. In some aspects, the overhang comprises a guide binding site that is at least partially complementary to a guide nucleic acid. In some aspects, the overhang comprises a genomic flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus. In some aspects, the donor strand comprises two overhangs, where the first overhang: comprises a first guide binding site that is at least partially complementary to a first guide nucleic acid; or a first genomic flap binding site that is at least partially identical or complementary to a first genomic flap at or adjacent to the genomic locus; and the second overhang: comprises a second guide binding site that is at least partially complementary to a second guide nucleic acid; or a second genomic flap binding site that is at least partially identical or complementary to a second genomic flap at or adjacent to the genomic locus. In some aspects, the donor strand corrects at least one genetic mutation in the at least one genomic locus. In some aspects, the donor strand comprises a coding sequence. In some aspects, the coding sequence encodes a full length protein or a fragment thereof. In some aspects, the donor strand comprises a non-coding sequence. In some aspects, the non-coding sequence knocks out an endogenous gene. In some aspects, the non-coding sequence comprises a regulatory element.

[00149] In some aspects, the system comprises a nuclease. The nuclease may be heterologous. In some aspects, the nuclease comprises an exonuclease for digesting the genomic flap. In some aspects, the exonuclease is a 5' exonuclease. Non-limiting example of the exonuclease can include a human flap endonuclease 1 (hFEN1), a human exonuclease 5 (hEXO5), a T5 exonuclease, a T7 exonuclease, an exonuclease VIII, a flap endonuclease domain of *E. coli* PolI, a RecJF, a Lambda exonuclease, a Xni (ExoIXI), a SaFEN (*Staphylococcus aureus* FEN), a nuclease BAL-31, or a fragment thereof. In some aspects, the exonuclease comprises an exonuclease in **Table 10**. In some aspects, the exonuclease comprises a polypeptide sequence at

least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the polypeptide sequence of any one of the exonuclease in **Table 10**.

Table 10. Non-limiting examples of exonuclease polypeptide sequence

Name	Exonuclease polypeptide sequence	SEQ ID NO:
hFEN1	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQF LIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMMENGIKPV YVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGAEQE VEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASC AALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKL PIQEFHLSRILQELGLNQEQFVDLCILLGSDYCESIRGIGPKR AVDLIQKHKSIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPE VLDPEVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLS KSRQGSTQGRLLDDFFKVTGSLSSAKRKEPEPKGSTKKKAKT GAAGKFKRGK	212
hFen1 (1-333)	MGIQGLAKLIADVAPSAIRENDIKSYFGRKVAIDASMSIYQF LIAVRQGGDVLQNEEGETTSHLMGMFYRTIRMMENGIKPV YVFDGKPPQLKSGELAKRSERRAEAEKQLQQAQAAGAEQE VEKFTKRLVKVTKQHNDECKHLLSLMGIPYLDAPSEAEASC AALVKAGKVYAAATEDMDCLTFGSPVLMRHLTASEAKKL PIQEFHLSRILQELGLNQEQFVDLCILLGSDYCESIRGIGPKR AVDLIQKHKSIEEIVRRLDPNKYPVPENWLHKEAHQLFLEPE VLDPEVELKWSEPNEEELIKFMCGEKQFSEERIRSGVKRLS KSRQ	213
hEXO5	MAETREEETVSAEASGFSDLSDSEFLEFLDLEDAQESKALV NMPGPSSESLGKDDKPISLQNWKRGLDILSPMERFHLKYL VTDLATQNWCELQTAYGKELPGFLAPEKAAVLDTGASIH ARELELHDLVTVPVTTKEDAWAIFLNILLIPTLQSEGHIRE FPVFGEGERVLLVGVIDELHYTAKGELELAELKTRRRPMLP LEAQKKKDCFQVSLYKYIFDAMVQGVTPASLIHHTKLCLE KPLGPSVLRHAQQGGFSVKSLGDLMELVFLSLTSLDLPVIDI LKIEYIHQETATVLGTEIVAFKEKEVRAKVQHMYMAYWMGH REPQGV DVEEAWKCRTCTYADICEWRKGSGLSSTLAPQV KKAK	214
T5 EXO	MSKSWGKFIEEEEAEMASRRNLMIVDGTNLGFRFKHNNSK KPFASSYVSTIQSLAKSYSARTTIVLGDGKGSVFRLEHLPEY KGNRDEKYAQRTEEEKALDEQFFEYLKDAFELCKTTFTFTI RGVEADDMAAYIVKLIGHLYDHVWLSTDGDWDTLTLDKV SRFSFTTRREYHLRDMYEHNVDDVEQFISLKAIMGDLGDN IRGVEGIGAKRGYNIREFGNVLDIIDQLPLPGKQKYIQNLNA SEELLFRNLILVDLPTYCVDAIAAVGQDVLDKFTKDILEIAE Q	215
T7 EXO	MALLDLKQFYELREGCDDKGILVMDGDWLVFQAMSAAEF DASWEEIWHRCCDHAKARQILEDSIKSYETRKKAWAGAPI VLAFTDSVNWRKELVDPNYKANRKAVKKPVGYFEFLDALF EREEFYCIREPMLEGDDVMGVASNPFAFGARKAVIISCDKD FKTIPNCDFLWCTTGNILTQTEESADWWHLFQTIKGDITDG YSGIAGWGDTAEDFLNNPFITEPKTSVLKSGKNKGQEVTKW VKRDPEPHETLWDCIKSIGAKAGMTEEDIKQGQMARILRF NEYNFIDKEIYLWRP	216

Name	Exonuclease polypeptide sequence	SEQ ID NO:
EXO VIII (RecE)	MSTKPLFLLRKAKKSSGEPDVVLWASNDFESTCATLDYLIV KSGKKLSSYFKAVATNFPVVDLPAEGEIDFTWSERYQLSK DSMTWELKPGAAPDNAHYQGNTNVNGEDMTEIEENMLLP SGQELPIRWLAQHGSEKPVTHVSRDGLQALHIARAEELPAV TALAVSHKTSLLDPLEIRELHKLVRDTRDKVFPNPGNSNLGLI TAFFEAYLNADYTDRGLLTKEWMKGNRVSHITRTASGANA GGGNLTDRGEGFVHDLTSLARDVATGVLARSMDLDIYNLH PAHAKRIEEIIAENKPPFSVFRDKFITMPGGLDYSRAIVVASV KEAPIGIEVIPAHVTEYLNKVLTTETDHANPDPEIVDIACGRSS APMPQRVTEEGKQDDEEKQPSGTTAVEQGEAETMEPDAT EHHQDTQPLDAQSQVNSVDAKYQELRAELHEARKNIPSKN PVDDDKLLAASRGEFVDGISDPNDPKWVKGIQTRDCVYQN QPETEKTSPPDMNQPEPVVQGEPEIACNACGQTGGDNCPCDG AVMGDATYQETFDDEESQVEAKENDPEEMEGAEPHNENA GSDPHRDCSDDETGEVADPVIVEDIEPGIYYGISNENYHAGPG ISKSQLDDIADTPALYLWRKNAPVDTTTKTKTLDLGTAFHCR VLEPEEFSNRFIVAPEFNRRTNAGKEEEKAFLEMECASTGKTV ITAEGRKIELMYQSVMALPLGQWLVESAGHAESSIYWEDP ETGILCRCPDKIPEFWIMDVKTTADIQRFKTAYDYRYH VQDAFYSDGYEAQFGVQPTFVFLVASTTIECGRYPVEIFMM GEEAKLAGQQEYHRNLRTLSDCLNTDEWPAIKTSLPRWA KEYAND	217
EXO VIII, truncated	EHPHNENAGSDPHRDCSDDETGEVADPVIVEDIEPGIYYGISN ENYHAGPGISKSQLDDIADTPALYLWRKNAPVDTTTKTKTLD LGTAFHCRVLEPEEFSNRFIVAPEFNRRTNAGKEEEKAFLEME CASTGKTVITAEGRKIELMYQSVMALPLGQWLVESAGHA ESSIYWEDPETGILCRCPDKIPEFWIMDVKTTADIQRFKT AYDYRYHVQDAFYSDGYEAQFGVQPTFVFLVASTTIECG RYPVEIFMMGEEAKLAGQQEYHRNLRTLSDCLNTDEWPAI KTLSLPRWAKEYAND	218
Flap endo domain of E coli PolI	VQIPQNPLILVDGSSYLRYAHAFPLTNSAGEPTGAMYGV LNMLRSLIMQYKPTHAAVVFDAKGKTRDELFEHYKSHRP PMPDDLRAQIEPLHAMVKAMGLPLAVSGVEADDVIGTLA REA EKAGRPVLISTGDKDMAQLVTPNITLINTMTNTILGPEE VVNKYGVPELIIDFLALMGDSSDNIPGVPGVGEKTAQALL QGLGGLDTLYAEPEKIAGLSFRGAKTMAAKLEQNKEVAYL SYQLATIKTDVELELTCEQLEVQQPAAEELLGLFKKYEFKR WTADVEAGKWLQAKGAKPAAKPQETSVADEAPEVTATVI	219
RecJ	MKQQIQLRRREVDETADLPAELPPLRLRYASRGVRS AQEL ERSVKGMLPWQQLSGVEKAVEILYNAFREGTRIIVVGDFDA DGATSTALSVLAMRSLGCSNIDYLVNRFEDGYGLSPEVVD QAHARGAQLIVTVDNGISSHAGVEHARSLGIPVIVTDHHL GDTLPAAEAIINPNLRDCNFPKSLAGVGVA FYLMLALRTFL RDQGWFDERNIAIPNLAELLDLVALGTVADVPLDANNRIL TWQGM SRIRAGKCRPGIKALLEVANRDAQKLAASDLGFAL GPRLNAAGRLDDMSVGVALLLCDNIGEARVLANELDALNQ TRKEIEQGMQIEALTLCEKLESRDTPGGLAMYHPEWHQG VVGILASRIKERFHRPVIAFAPAGDGT LKSGRSIQGLHMRD ALERLDTLYPGMMLKFGGHAMAAGLSLEEDKFKLFQQRFG ELVTEWLDPSLLQGEVVS DGPLSPAEMTMEVAQLLRDAGP	220

Name	Exonuclease polypeptide sequence	SEQ ID NO:
	WGQMFPEPLFDGHFRLQQLVGERHLKVMVEPVGGGPLL DGIAFNVDALWPDNGVREVQLAYKLDINEFRGNRSLQIIID NIWPI	
Lambda exo	MTPDIILQRTGIDVRAVEQGDDAWHKLRLGVITASEVHNV AKPRSGKKWPDMSYFHTLLAEVCTGVAPEVNAKALAW GKQYENDARTLFEFTSGVNVTESPIIYRDESMRTACSPDGLC SDGNGLKCPFTSRDFMKFRLGGFEAIKSAYMAQVQYSM WVTRKNAWYFANYDPRMKREGLHYVVIERDEKYMASFDE IVPEFIEKMDEALAEIGFVFGEQWR	154
Xni (ExoIXI) from E coli	MAVHLLIVDALNLIRRIHAVQGSPCVETCQHALDQLIMHSQ PTHAVAVFDDENRSSGWRHQRLPDYKAGRPPMPEELHDEM PALRAAFEQRGVPCWSTSGNEADDLAATLAVKVTQAGHQ ATIVSTDKGYCQLSPTLRIRDYFQKRWLDAPFIDKEFGVQP QQLPDYWGLAGISSKVPVAGIGPKSATQLLVEFQSLEGIY ENLDAVAEKWRKKLETHKEMAFLCRDIARLQTDLHIDGNL QQLRLVR	155
SaFEN (Staph aureus)	MPNKILLVDGMALLFRHFYATSLHKQFMYSQGVPTNGIQ GFVRHIFSAIHEIRPTHVAVCWDMGQSTFRNDMFDGYKQN RSAPPEELIPQFDYVKEISEQFGFVNIGVKNYEADDVIGTLA QQYSTDNVYIITGDKDLLQCINDNVEVWLIKKGFNIYNRY TLHRFNEEYALEPQQLIDIKAFMGDTADGYAGVKGIGEKTA IKLIQYQSVENVVENIDALSAGQRNKINDNLDELYLSKRL AEIHTQVPIDSEALFEKMSFATTLNHILSICNEHELHVSGKYI SSHF	156

[00150] In some aspects, the system comprises at least one additional endonuclease that is different from the at least one programmable endonuclease described herein. In some aspects, the at least one additional endonuclease can digest the genomic flap.

[00151] In some aspects, the system comprises a dominant negative MMR peptide to improve genomic editing capability, particularly in cells which overexpress the MMR pathway. In some aspects, the dominant negative MMR peptide can be delivered as a fusion (e.g., fused with any component of the system described herein), recruited, or as separate peptide. **Table 11** lists non-limiting examples of the MMR peptide sequences.

Table 11. Non-limiting examples of MMR polypeptide sequence

Name	MMR peptide sequence	SEQ ID NO:
MLH 1	MSFVAGVIRRLDETVVNRIAAGEVIQRPANA KEMIENCLDAKSTSIQVI VKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEDLASISTYGFR GEALASISHVAHVTTITTKTADGKCAYRASYS DGKLKAPPKPCAGNQGT QITVEDLFYNIATRRKALKNPSEEGKILEVVGRYSVHNAGISFSVKKQ GETVADVRTLPLNASTVDNIRSIFGNAVSR ELIEIGCEDKTLAFKMNGYIS NANYSVKKKCIFFLLFINHRLVESTSLRKA IETVYAAAYLPKNTHPFLYLSLE ISPQNVVDVNVHPTKHEVHFLHEESILERVQQHIESKLLGSNSSRMYFTQT LLPGLAGPSGEMVKSTTSLTSSSTSGSSDKVYAHQMVRTDSREQKLDA FLQPLSKPLSSQPQAIVTEDKTDISSGRARQQDEEMLELPAPAEVAAKN QSLEGDTTKGTSEMSEKRGPTSSNPRKRHREDS D VEMVEDDSRKEMTA ACTPRRRIINLTSVLSLQEEINEQGHEVLREMLHNHSFVGC VNPQWALA QHQTCLYLLNTTKLSEELFYQILYDFANFGVLRLSEPAPLFDLAMLAL DSPESGWTEEDGPKEGLAEYIVEFLKKKAEM LADYFSLEIDEEGNLIGL PLLIDNYVPPLEGLPIFILRLATEVNWDEEKECFESLSKECAMFY SIRQ YISEESTLSGQQSEVPGSIPNSWKWTVEHIVYKALRSHILPPKHFTEDGN ILQLANLPDLYKVFERC	157
MLH 1 E34A	MSFVAGVIRRLDETVVNRIAAGEVIQRPANA KAMIENCLDAKSTSIQVI VKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEDLASISTYGFR GEALASISHVAHVTTITTKTADGKCAYRASYS DGKLKAPPKPCAGNQGT QITVEDLFYNIATRRKALKNPSEEGKILEVVGRYSVHNAGISFSVKKQ GETVADVRTLPLNASTVDNIRSIFGNAVSR ELIEIGCEDKTLAFKMNGYIS NANYSVKKKCIFFLLFINHRLVESTSLRKA IETVYAAAYLPKNTHPFLYLSLE ISPQNVVDVNVHPTKHEVHFLHEESILERVQQHIESKLLGSNSSRMYFTQT LLPGLAGPSGEMVKSTTSLTSSSTSGSSDKVYAHQMVRTDSREQKLDA FLQPLSKPLSSQPQAIVTEDKTDISSGRARQQDEEMLELPAPAEVAAKN QSLEGDTTKGTSEMSEKRGPTSSNPRKRHREDS D VEMVEDDSRKEMTA ACTPRRRIINLTSVLSLQEEINEQGHEVLREMLHNHSFVGC VNPQWALA QHQTCLYLLNTTKLSEELFYQILYDFANFGVLRLSEPAPLFDLAMLAL DSPESGWTEEDGPKEGLAEYIVEFLKKKAEM LADYFSLEIDEEGNLIGL PLLIDNYVPPLEGLPIFILRLATEVNWDEEKECFESLSKECAMFY SIRQ YISEESTLSGQQSEVPGSIPNSWKWTVEHIVYKALRSHILPPKHFTEDGN ILQLANLPDLYKVFERC	158
MLH 1 del75 4-756	MSFVAGVIRRLDETVVNRIAAGEVIQRPANA KEMIENCLDAKSTSIQVI VKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEDLASISTYGFR GEALASISHVAHVTTITTKTADGKCAYRASYS DGKLKAPPKPCAGNQGT QITVEDLFYNIATRRKALKNPSEEGKILEVVGRYSVHNAGISFSVKKQ GETVADVRTLPLNASTVDNIRSIFGNAVSR ELIEIGCEDKTLAFKMNGYIS NANYSVKKKCIFFLLFINHRLVESTSLRKA IETVYAAAYLPKNTHPFLYLSLE ISPQNVVDVNVHPTKHEVHFLHEESILERVQQHIESKLLGSNSSRMYFTQT LLPGLAGPSGEMVKSTTSLTSSSTSGSSDKVYAHQMVRTDSREQKLDA FLQPLSKPLSSQPQAIVTEDKTDISSGRARQQDEEMLELPAPAEVAAKN QSLEGDTTKGTSEMSEKRGPTSSNPRKRHREDS D VEMVEDDSRKEMTA ACTPRRRIINLTSVLSLQEEINEQGHEVLREMLHNHSFVGC VNPQWALA QHQTCLYLLNTTKLSEELFYQILYDFANFGVLRLSEPAPLFDLAMLAL DSPESGWTEEDGPKEGLAEYIVEFLKKKAEM LADYFSLEIDEEGNLIGL PLLIDNYVPPLEGLPIFILRLATEVNWDEEKECFESLSKECAMFY SIRQ	159

Name	MMR peptide sequence	SEQ ID NO:
	YISEESTLSGQQSEVPGSIPNSWKWTVEHIVYKALRSHILPPKHFTEDGN ILQLANLPDLYKVF	
MLH 1 E34A del 754-756	MSFVAGVIRRLDETVVNRIAAGEVIQRPANAIAKAMIENCLDAKSTSIQVI VKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEDLASISTYGFR GEALASISHVAHVTTITTKTADGKCAYRASYSYGKLGKAPPKPCAGNQGT QITVEDLFYNIATRRKALKNPSEEGKILEVVGRYSVHNAGISFSVKKQ GETVADVRTLTPNASTVDNIRSIFGNAVSRLEIEIGCEDKTLAFKMNGYIS NANYSVKKKCIFFLLFINHRLVESTSLRKAJETVYAAAYLPKNTHPFLYLSLE ISPQNVVDVNVHPTKHEVHFLHEESILERVQQHIESKLLGSNSSRMYFTQT LLPGLAGPSGEMVKSTTSLTSSSTSGSSDKVYAHQMVRTDSREQKLDA FLQPLSKPLSSQPQAIVTEDKTDISSGRARQQDEEMLELPAPAEVAANK QSLEGDTTKGTSEMSEKRGPTSSNPRKRHRESDVEMVEDDSRKEMTA ACTPRRRIINLTSVLSLQEEINEQGHEVLREMLHNHSFVGCVPNPQWALA QHQTCLYLLNTTKLSEELFYQILIYDFANFGVLRLSEPAPLFDLAMLAL DSPESGWTEEDGPKEGLAEYIVEFLKKKAEMLADYFSLEIDEEGNLIGL PLLIDNYVPPLEGLPIFILRLATEVNWDEEKECFESLSKECAMFY SIRQQ YISEESTLSGQQSEVPGSIPNSWKWTVEHIVYKALRSHILPPKHFTEDGN ILQLANLPDLYKVF	160
MLH 1 1-335	MSFVAGVIRRLDETVVNRIAAGEVIQRPANAIAKEMIENCLDAKSTSIQVI VKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEDLASISTYGFR GEALASISHVAHVTTITTKTADGKCAYRASYSYGKLGKAPPKPCAGNQGT QITVEDLFYNIATRRKALKNPSEEGKILEVVGRYSVHNAGISFSVKKQ GETVADVRTLTPNASTVDNIRSIFGNAVSRLEIEIGCEDKTLAFKMNGYIS NANYSVKKKCIFFLLFINHRLVESTSLRKAJETVYAAAYLPKNTHPFLYLSLE ISPQNVVDVNVHPTKHEVHFLHEESILERVQQHIESKLL	161
MLH 1 1-335 E34A	MSFVAGVIRRLDETVVNRIAAGEVIQRPANAIAKAMIENCLDAKSTSIQVI VKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEDLASISTYGFR GEALASISHVAHVTTITTKTADGKCAYRASYSYGKLGKAPPKPCAGNQGT QITVEDLFYNIATRRKALKNPSEEGKILEVVGRYSVHNAGISFSVKKQ GETVADVRTLTPNASTVDNIRSIFGNAVSRLEIEIGCEDKTLAFKMNGYIS NANYSVKKKCIFFLLFINHRLVESTSLRKAJETVYAAAYLPKNTHPFLYLSLE ISPQNVVDVNVHPTKHEVHFLHEESILERVQQHIESKLL	162
MSH2	MAVQPKETLQLESAAEVGFVRFFQGMPEKPTTTVRLFDRGDFYTAHGE DALLAAREVFKTQGVIKYMGPAKAKNLQSVVLSKMNFESEFVKDLLLV RQYRVEVYKNRAGNKASKENDWYLAYKASPGNLSQFEDILFGNNDMS ASIGVVGVKMSAVDQGQRQVGVGYVDSIQRKLGLCEFPDNDQFSNLEA LLIQIGPKECVLPGGETAGDMGKLRQIIQRGGILITERKKADFSTKDIYQ DLNRLKGGKGEQMNSAVLPEMENQVAVSSLSAVIKFLELLSDDSNFG QFELTTDFDSQYMKLDIAAVRALNLFQGSVEDTTGSQSLAALLNKCKTP QGQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDLLRRF PDLNRLAKKFQRQAANLQDCYRLYQGINQLPNVIALEKHEGKHQKL LLAVFVTPLTDLRSDFSKFQEMIETTLMDMDQVENHEFLVKPSFDPNLSE LREIMNDLEKKMQSTLISAARDLGLDPGKQIKLDSSAQFGYYFRVTCKE	163

Name	MMR peptide sequence	SEQ ID NO:
	EKVLRNNKNFSTVDIQKNGVKFTNSKLTSLNEEYTKNKTEYEEAQDAI VKEIVNISSGYVEPMQTLNDVLAQLDAVVSFAHVSNGAPVPYVRPAILE KGQGRILKASRHACVEVQDEIAFIPNDVYFEKDKQMFHIITGPNMGGK STYIRQTGVIVLMAQIGCFVPCESA EVSIVDCILARVGAGDSQLKGVSTF MAEMLETASILRSATKDSLIIIDELGRGTSTYDGFGLAWAISEYIATKIGA FCMFATHFHELTALANQIPTVNNLHVTALTTEETLTMLYQVKKGVCDQ SFGIHVAELANFPKHVIECAKQKALELEEFQYIGESQGYDIMEPAAKKC YLEREQGEKIIQEFLSKVKQMPFTEMSEENITIKLKQLKAEVIAKNNSFV NEIISRIKVT	
MSH2 G674 A	MAVQPKETLQLESAAEVGFVRFFQGMPEKPTTTVRLFDRGDFYTAHGE DALLAAREVFKTQGVYKYMGPAGAKNLQSVVLSKMNFESEFVKDLLLV RQYRVEVYKNRAGNKASKENDWYLAYKASPGNLSQFEDILFGNNDMS ASIGVVGKMSAVDGGQRQVGVGYSIQRKLGLCEFPDNDQFSNLEA LLIQIGPKECVLPGETAGDMGKLRQIIQRGGILITERKKADFSTKDIYQ DLNRLKGGKGEQMNSAVLPEMENQVAVSSLSAVIKFLELLSDDSNFG QFELTTDFDSQYMKLDIAAVRALNLFQGSVEDTTGSQSLAALLNKCKTP QGQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDLLRRF PDLNRLAKKFQRQAANLQDCYRLYQGINQLPNVIALEKHEGKHQKL LLAVFVTPLTDLRSDFSQFQEMETTLMDMDQVENHEFLVKPSFDPNLSE LREIMNDLEKKMQSTLISAARDLGLDPGKQIKLDSSAQFGYYFRVTCKE EKVLRNNKNFSTVDIQKNGVKFTNSKLTSLNEEYTKNKTEYEEAQDAI VKEIVNISSGYVEPMQTLNDVLAQLDAVVSFAHVSNGAPVPYVRPAILE KGQGRILKASRHACVEVQDEIAFIPNDVYFEKDKQMFHIITGPNMGAK STYIRQTGVIVLMAQIGCFVPCESA EVSIVDCILARVGAGDSQLKGVSTF MAEMLETASILRSATKDSLIIIDELGRGTSTYDGFGLAWAISEYIATKIGA FCMFATHFHELTALANQIPTVNNLHVTALTTEETLTMLYQVKKGVCDQ SFGIHVAELANFPKHVIECAKQKALELEEFQYIGESQGYDIMEPAAKKC YLEREQGEKIIQEFLSKVKQMPFTEMSEENITIKLKQLKAEVIAKNNSFV NEIISRIKVT	164
MSH2 N671I	MAVQPKETLQLESAAEVGFVRFFQGMPEKPTTTVRLFDRGDFYTAHGE DALLAAREVFKTQGVYKYMGPAGAKNLQSVVLSKMNFESEFVKDLLLV RQYRVEVYKNRAGNKASKENDWYLAYKASPGNLSQFEDILFGNNDMS ASIGVVGKMSAVDGGQRQVGVGYSIQRKLGLCEFPDNDQFSNLEA LLIQIGPKECVLPGETAGDMGKLRQIIQRGGILITERKKADFSTKDIYQ DLNRLKGGKGEQMNSAVLPEMENQVAVSSLSAVIKFLELLSDDSNFG QFELTTDFDSQYMKLDIAAVRALNLFQGSVEDTTGSQSLAALLNKCKTP QGQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDLLRRF PDLNRLAKKFQRQAANLQDCYRLYQGINQLPNVIALEKHEGKHQKL LLAVFVTPLTDLRSDFSQFQEMETTLMDMDQVENHEFLVKPSFDPNLSE LREIMNDLEKKMQSTLISAARDLGLDPGKQIKLDSSAQFGYYFRVTCKE EKVLRNNKNFSTVDIQKNGVKFTNSKLTSLNEEYTKNKTEYEEAQDAI VKEIVNISSGYVEPMQTLNDVLAQLDAVVSFAHVSNGAPVPYVRPAILE KGQGRILKASRHACVEVQDEIAFIPNDVYFEKDKQMFHIITGPIIMGGKS TYIRQTGVIVLMAQIGCFVPCESA EVSIVDCILARVGAGDSQLKGVSTF MAEMLETASILRSATKDSLIIIDELGRGTSTYDGFGLAWAISEYIATKIGA FCMFATHFHELTALANQIPTVNNLHVTALTTEETLTMLYQVKKGVCDQ SFGIHVAELANFPKHVIECAKQKALELEEFQYIGESQGYDIMEPAAKKC YLEREQGEKIIQEFLSKVKQMPFTEMSEENITIKLKQLKAEVIAKNNSFV NEIISRIKVT	165

[00152] The system may relate to a 1-sided Replacer 1. Some aspects include a system comprising: (a) at least one RNA-guided endonuclease; (b) at least one guide nucleic acid comprising: (i) a spacer complementary to a genomic locus in a cell, (ii) a scaffold for complexing with the at least one RNA-guided endonuclease, (iii) an optional donor binding site that is at least partially complementary to an integrating nucleic acid, and (iv) a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus; and (c) at least one DNA ligase; and (d) the integrating nucleic acid, optionally comprising a guide binding site that is at least partially complementary to the at least one guide nucleic acid, wherein the at least one RNA-guided endonuclease cleaves at least one strand of the genomic locus, and wherein the at least one DNA ligase ligates an end of the integrating nucleic acid to the genomic flap site, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. The integrating nucleic acid may comprise a single-stranded DNA.

[00153] The system may relate to a 2-sided Replacer 1. Some aspects include a system comprising: (a) at least one RNA-guided endonuclease comprising a first RNA-guided endonuclease and an optional second RNA-guided endonuclease; (b) at least one guide nucleic acid comprising a first guide nucleic acid and a second guide nucleic acid, the first guide nucleic acid comprising: (i) a first spacer complementary to a first region of a genomic locus in a cell, (ii) a first scaffold for complexing with the first RNA-guided endonuclease, and (iii) an optional first donor binding site that at least partially complementary to an integrating nucleic acid, and (iv) a first flap binding site that is at least partially identical or complementary to a first genomic flap at or adjacent to the genomic locus; and the second guide nucleic acid comprising: (i) a second spacer complementary to a second region of the genomic locus in the cell, (ii) a second scaffold for complexing with the first or second RNA-guided endonuclease, (iii) an optional second donor binding site that at least partially complementary to the integrating nucleic acid, and (iv) a second flap binding site that is at least partially identical or complementary to a second genomic flap at or adjacent to the genomic locus; (c) at least one DNA ligase comprising a first DNA ligase and an optional second DNA ligase; and (d) at least one integrating nucleic acid comprising a first strand and a second strand: (i) wherein the first strand comprises an optional first guide binding site that is at least partially complementary to the first guide nucleic acid, and (ii) wherein the second strand comprises an optional second guide binding site that is at least partially complementary to the second guide nucleic acid, wherein the first RNA-guided endonuclease and/or the second RNA-guided endonuclease each cleaves at least one strand of the genomic locus in the cell; and wherein the first DNA ligase ligates an end of the first strand of the integrating nucleic acid to the first genomic flap; and the

first or second DNA ligase ligates an end of the second strand of the integrating nucleic acid to the second genomic flap, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. The integrating nucleic acid may comprise a double-stranded DNA duplex region. The integrating nucleic acid may comprise a 5' overhang optionally comprising the first guide binding site. The integrating nucleic acid may comprise a 5' overhang optionally comprising the second guide binding site.

[00154] The system may relate to 1-sided Replacer 2. Some aspects include a system comprising: (a) at least one RNA-guided endonuclease; (b) at least one guide nucleic acid comprising: (i) a spacer complementary to a genomic locus in a cell, (ii) a scaffold for complexing with the at least one RNA-guided endonuclease, and (iii) an optional donor binding site that is at least partially complementary to an integrating nucleic acid; (c) at least one DNA ligase; and (d) the integrating nucleic acid that: (i) comprises an optional guide binding site that is at least partially complementary to the at least one guide nucleic acid, and (ii) comprises a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus, wherein the at least one RNA-guided endonuclease cleaves at least one strand of the genomic locus; and wherein the at least one DNA ligase ligates an end of the integrating nucleic acid to the genomic flap, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. The integrating nucleic acid may comprise a DNA comprising a 3' overhang. The 3' overhang may comprise the guide binding site. The 3' overhang may comprise the flap binding site. The at least one DNA ligase may ligates a strand of the integrating nucleic acid to the genomic nucleic acid sequence.

[00155] The system may relate to 2-sided Replacer 2. Some aspects include a system comprising: (a) at least one RNA-guided endonuclease comprising a first RNA-guided endonuclease and an optional second RNA-guided endonuclease; (b) at least one guide nucleic acid comprising a first guide nucleic acid and a second guide nucleic acid, the first guide nucleic acid comprising: (i) a first spacer complementary to a first region of a genomic locus in a cell, (ii) a first scaffold for complexing with the first RNA-guided endonuclease, and (iii) an optional first donor binding site that at least partially complementary to an integrating nucleic acid; and the second guide nucleic acid comprising: (i) a second spacer complementary to a second region of the genomic locus in the cell, (ii) a second scaffold for complexing with the first or second RNA-guided endonuclease, and (iii) an optional second donor binding site that at least partially complementary to the integrating nucleic acid; and at least one DNA ligase comprising a first DNA ligase and an optional second DNA ligase; and the integrating nucleic acid comprising a first strand and a second strand: wherein the first strand comprises an optional first guide binding site that is at least partially complementary to the first guide nucleic

acid; wherein the second strand comprises an optional second binding site that is at least partially complementary to the second guide nucleic acid; wherein the first strand comprises a first flap binding site that is at least partially identical or complementary to a first genomic flap at or adjacent to the genomic locus; and wherein the second strand comprises a second flap binding site that is at least partially identical or complementary to a second genomic flap at or adjacent to the genomic locus; wherein the first RNA-guided endonuclease and/or the second RNA-guided endonuclease each cleaves at least one strand of the genomic locus in the cell; and wherein the first DNA ligase ligates an end of the first strand of the integrating nucleic acid to the first genomic flap; and the first or second DNA ligase ligates an end of the second strand of the integrating nucleic acid to the second genomic flap, thereby replacing a region of the genomic locus with the integrating nucleic acid in the cell. The integrating nucleic acid may comprise a double-stranded DNA duplex region. The double-stranded DNA may comprise a 3' overhang optionally comprising the first guide binding site, and comprising the first flap binding site. The double stranded DNA may comprise a 3' overhang optionally comprising the second guide binding site, and comprising the second flap binding site.

[00156] In the system, the at least one RNA-guided endonuclease may comprise a Cas protein or a functional fragment thereof. The Cas protein or the functional fragment thereof may comprise nickase activity. The at least one RNA-guided endonuclease may comprise a Cas9 nickase or a functional fragment thereof. The at least one DNA ligase may ligates nucleic acids bound to DNA. The at least one DNA ligase may ligates nucleic acids bound to RNA. The at least one DNA ligase may comprise a PBCV-1 DNA ligase. The at least one DNA ligase may be operatively coupled to the at least one RNA-guided endonuclease. The at least one DNA ligase may be fused to the at least one RNA-guided endonuclease as a fusion polypeptide. The at least one RNA-guided endonuclease and the at least one DNA ligase may comprise a heterodimer domain. The at least one RNA-guided endonuclease and the at least one DNA ligase may form a heterodimer via the heterodimer domain. The at least one RNA-guided endonuclease may comprise a linker. The linker may connect the Cas protein or a functional fragment thereof to the heterodimer domain. The at least one RNA-guided endonuclease may comprise a localization signal sequence. The at least one DNA ligase may comprise a localization signal sequence. The localization signal sequence may comprise a nuclear localization sequence (NLS). The at least one RNA-guided endonuclease or the at least one DNA ligase may be directed to nucleus of the cell by the NLS. The at least one integrating nucleic acid may correct at least one genetic mutation in the at least one genomic locus. The at least one integrating nucleic acid may insert a coding sequence. The coding sequence may encode a full length protein. The at least one integrating nucleic acid may insert a non-coding

sequence. The non-coding sequence may knock out an endogenous gene. The non-coding sequence may comprise a regulatory element. The system may further include a nuclease. The nuclease may comprise an exonuclease for digesting the genomic flap. The nuclease may comprise a human flap endonuclease 1 (hFEN1), a human exonuclease 5 (hEXO5), a T5 exonuclease, a T7 exonuclease, an exonuclease VIII, a flap endonuclease domain of *E. coli* PolI, a RecJF, a Lambda exonuclease, a Xni (ExoIXI), a SaFEN (Staphylococcus aureus FEN), a nuclease BAL-31, or a fragment thereof. The heterologous nuclease may comprise an endonuclease for digesting the genomic flap, and the endonuclease may be different from the at least one RNA-guided endonuclease. The at least one RNA-guided endonuclease may comprise at least one additional functional domain. The at least one additional functional domain may comprise a chromatin modifying domain. The at least one additional functional domain may comprise a cell penetrating peptide. The at least one guide nucleic acid may comprise at least one nucleic acid modification. The at least one nucleic acid modification may comprise a modification to a backbone, a sugar, a base, or a combination thereof. The at least one RNA-guided endonuclease may be complexed with the at least one guide nucleic acid. The at least one guide nucleic acid may be complexed with the integrating nucleic acid. The at least one RNA-guided endonuclease, the at least one guide nucleic acid, the at least one at least one DNA ligase, the integrating nucleic acid, or a combination thereof may be encoded by a polynucleotide. The polynucleotide may comprise mRNA. The polynucleotide may comprise a vector. The vector may comprise a viral vector. The at least one RNA-guided endonuclease, the at least one guide nucleic acid, the at least one at least one DNA ligase, the integrating nucleic acid, or a combination thereof may be encapsulated by at least one lipid nanoparticle. The cell may comprise a bacterial cell or a prokaryotic cell. The cell may include a prokaryotic cell. The prokaryotic cell may include a bacterial cell. The editing may be performed in a cytoplasm of the bacterial cell. The cell may include a eukaryotic cell. The eukaryotic cell may include an animal cell or a plant cell. The eukaryotic cell may include a plant cell. The eukaryotic cell may include an animal cell. The eukaryotic cell may comprise a mammalian cell. The editing may be performed in a cytoplasm of the eukaryotic cell. The editing may be performed in a nucleus of the eukaryotic cell. The system, or any aspect of the system, may be included in a composition, or in a cell such as a cell line.

[00157] Some aspects relate to a system that includes nucleic acids. The system may include guide nucleic acids, integrating nucleic acids, or a combination thereof. Some aspects relate to a system of nucleic acids. The system may include a system of guide nucleic acids. The system may include a system of integrating nucleic acids. The system of nucleic acids may further include other aspects such as additional nucleic acids or non-nucleic acid components.

[00158] The system of nucleic acids may include a guide nucleic acid. The guide nucleic acid may include a spacer. The spacer may be complementary to a region of a locus (e.g. genomic locus) of a target nucleic acid such as a genomic strand. The target nucleic acid may be in a cell. The genomic strand may be in a cell. The target nucleic acid may be *in vitro*. The guide nucleic acid may include a scaffold. The scaffold may complex with an endonuclease such as an RNA-guided endonuclease. The guide nucleic acid may include a flap binding site. The flap binding site may be complementary or at least partially complementary to a flap such as a genomic flap. The flap binding site may be identical or at least partially identical to a flap such as a genomic flap. The flap may be at the locus. The flap may be adjacent to the locus. The guide nucleic acid may include a donor binding site. The donor binding site may be complementary to an integrating nucleic acid. The donor binding site may be partially complementary to an integrating nucleic acid. The donor binding site may be complementary to a splinting nucleic acid. The donor binding site may be partially complementary to a splinting nucleic acid. Components of the guide nucleic acid may be included in 1 guide nucleic acid. More than one guide nucleic acid may be used. Components of the guide nucleic acid may collectively be included among multiple guide nucleic acids. Components of the guide nucleic acid may split between multiple guide nucleic acids.

[00159] The system of nucleic acids may include an integrating nucleic acid. The integrating nucleic acid may include a 5' end to be ligated. The 5' end may be ligated. The 5' end may be ligated to a 3' terminus. The 3' terminus may be of a target nucleic acid strand (e.g. a genomic strand). The 3' terminus may be generated by an endonuclease such as an RNA-guided endonuclease. The integrating nucleic acid may include a 5' end to be ligated to a 3' terminus of a genomic strand generated by an RNA-guided endonuclease. Components of the integrating nucleic acid may be included in 1 or 2 complementary strands. Components of the integrating nucleic acid may be included in 1 integrating nucleic acid. More than one integrating nucleic acid may be used. Components of the integrating nucleic acid may collectively be included among multiple integrating nucleic acids. Components of the integrating nucleic acid may split between multiple integrating nucleic acids.

[00160] The system of nucleic acids may include a splinting nucleic acid (also referred to as a "splinting strand"). The splinting strand may hybridize to two nucleic acids comprising ends to be ligated. The splinting nucleic acid may include a flap binding site. The flap binding site may be complementary to a flap. The flap binding site may be partially complementary to a flap. The flap binding site may be identical to a flap. The flap binding site may be partially identical to a flap. The flap may be at a locus of a target nucleic acid. The flap may be adjacent to a locus of a target nucleic acid. The flap may be a genomic flap. The locus may be a genomic locus.

The flap binding site may be at least partially identical or complementary to a genomic flap at or adjacent to a genomic locus. The splinting nucleic acid may include a guide binding site. The guide binding site may be complementary to a guide nucleic acid. The guide binding site may be partially complementary to a guide nucleic acid. Components of the splinting nucleic acid may be included in 1 splinting nucleic acid. More than one splinting nucleic acid may be used. The splinting nucleic acid may include a donor binding site. The donor binding site may be complementary to an integrating nucleic acid. The donor binding site may be partially complementary to an integrating nucleic acid.

[00161] The splinting strand may be or include DNA. The splinting strand may be or include RNA. The splinting nucleic acid may be included as part of an integrating nucleic acid. The splinting nucleic acid may be included as a strand of a double stranded integrating nucleic acid. The splinting nucleic acid may be included as part of a guide nucleic acid.

[00162] The system of nucleic acids may include: (a) a guide nucleic acid comprising: (i) a spacer complementary to a region of a genomic locus of a genomic strand, (ii) a scaffold for complexing with RNA-guided endonuclease, (iii) an optional donor binding site that is at least partially complementary to an integrating nucleic acid, and (iv) a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus; and (b) an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by an RNA-guided endonuclease. A component of (i), (ii), (iii), or (iv) may be included in a single guide nucleic acid, or may be split between or collectively included among multiple guide nucleic acids.

[00163] The system of nucleic acids may include: (a) a guide nucleic acid comprising (i) a spacer complementary to a region of a genomic locus of a genomic strand, (ii) a scaffold for complexing with an RNA-guided endonuclease, and (iii) an optional donor binding site that is at least partially complementary to a splinting nucleic acid; (b) an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by an RNA-guided endonuclease; and (c) a splinting nucleic acid comprising a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus, and comprising an optional guide binding site that is at least partially complementary to a guide nucleic acid. A component of (i), (ii), or (iii) may be included in a single guide nucleic acid, or may be split between or collectively included among multiple guide nucleic acids.

[00164] In some aspects, the system described herein can be delivered into a cell, where one or more of the components of the system can be delivered into the cell together. In some aspects, each component of the system can be delivered into the cell separately. In some aspects, the system can be encoded by a polynucleotide such as a heterologous polynucleotide, where the

polynucleotide is delivered into a cell and where the polynucleotide is expressed by the cell to generate the components of the cell. In some aspects, the system can be encoded and delivered into the cell via a polynucleotide comprising mRNA. In some aspects, the system can be encoded and delivered into the cell via a polynucleotide comprising a vector. In some aspects, the vector comprises a viral vector. The system can be encapsulated in a lipid or nanoparticle, or multiple lipids or nanoparticles. In some aspects, the system can be encapsulated in at least one lipid nanoparticle. In some aspects, the system comprises a ribonucleoprotein (RNP). For example, at least one RNA-guided endonuclease described herein (e.g., a Cas9) can be complexed with at least one guide nucleic acid described herein (e.g., forming a CRISPR ribonucleoprotein) for delivery. In some aspects, the system comprises at least one RNP comprising a RNA-guided endonuclease complexed with at least one first guide nucleic acid or with at least one second guide nucleic acid. In some aspects, the system comprises at least one RNP and at least one integrating nucleic acid (e.g., a single-stranded or a double-stranded integrating nucleic acid described herein). In some aspects, the system comprises at least one RNP and at least one integrating nucleic acid. In some aspects, the system comprises at least one RNP and at least one first integrating nucleic acid or at least one second integrating nucleic acid.

[00165] In some aspects, the system described herein can modify a genomic locus or gene in a cell. In some aspects, the cell comprises a bacterial cell, an eukaryotic cell, or a plant cell. In some aspects, the system described herein can be formulated into a composition, a pharmaceutical composition, a kit, or a combination thereof. In some aspects, the system described herein can be delivered and propagated in a cell line.

[00166] Some aspects include an editing system, comprising an RNA-guided endonuclease, a guide nucleic acid, and an integrating nucleic acid. Some aspects include an editing method, comprising: contacting a target nucleic acid with the editing system and a DNA ligase.

Pharmaceutical compositions

[00167] Described herein, in some aspects, is a pharmaceutical composition comprising the system or the composition described herein. The pharmaceutical composition may include a pharmaceutically acceptable excipient, carrier, or diluent. The pharmaceutical composition may include a carrier. The pharmaceutical composition may include an excipient. The pharmaceutical composition may be delivered to a subject. The pharmaceutical composition may be delivered to a cell. The pharmaceutical composition may be used in a method disclosed herein.

[00168] The pharmaceutical compositions described herein comprise the system, the composition, or the cell contacted with the system or contacted with the composition. The

pharmaceutical composition may comprise a composition such as a protein or nucleic acid disclosed herein. The pharmaceutical composition may comprise a cell comprising a composition or system disclosed herein.

[00169] A pharmaceutical composition may include a mixture of a pharmaceutical composition, with other chemical components (i.e. pharmaceutically acceptable inactive ingredients), such as carriers, excipients, binders, filling agents, suspending agents, flavoring agents, sweetening agents, disintegrating agents, dispersing agents, surfactants, lubricants, colorants, diluents, solubilizers, moistening agents, plasticizers, stabilizers, penetration enhancers, wetting agents, anti-foaming agents, antioxidants, preservatives, or one or more combination thereof. In practicing the methods of treatment or use provided herein, therapeutically effective amounts of pharmaceutical compositions described herein are administered to a mammal having a disease, disorder, or condition to be treated. In some aspects, the mammal is a human. A therapeutically effective amount can vary widely depending on the severity of the disease, the age and relative health of the subject, the potency of the pharmaceutical composition used and other factors. The pharmaceutical compositions can be used singly or in combination with one or more pharmaceutical compositions as components of mixtures.

[00170] The pharmaceutical composition may be formulated for administering intrathecally, intraocularly, intravitreally, retinally, intravenously, intramuscularly, intraventricularly, intracerebrally, intracerebellarly, intracerebroventricularly, intraparenchymally, subcutaneously, intratumorally, pulmonarily, endotracheally, intraperitoneally, intravesically, intravaginally, intrarectally, orally, sublingually, transdermally, by inhalation, by inhaled nebulized form, by intraluminal-GI route, or a combination thereof to a subject in need thereof.

[00171] The pharmaceutical formulations described herein are administered to a subject by appropriate administration routes, including but not limited to, intravenous, intraarterial, oral, parenteral, buccal, topical, transdermal, rectal, intramuscular, subcutaneous, intraosseous, transmucosal, inhalation, or intraperitoneal administration routes. The pharmaceutical formulations described herein include, but are not limited to, aqueous liquid dispersions, self-emulsifying dispersions, solid solutions, liposomal dispersions, aerosols, solid dosage forms, powders, immediate release formulations, controlled release formulations, fast melt formulations, tablets, capsules, pills, delayed release formulations, extended release formulations, pulsatile release formulations, multiparticulate formulations, and mixed immediate and controlled release formulations. Pharmaceutical compositions including a pharmaceutical composition are manufactured in a conventional manner, such as, by way of example only, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or compression processes.

[00172] The pharmaceutical compositions may include at least a pharmaceutical composition as an active ingredient in free-acid or free-base form, or in a pharmaceutically acceptable salt form. In addition, the methods and pharmaceutical compositions described herein include the use of N-oxides (if appropriate), crystalline forms, amorphous phases, as well as active metabolites of these compounds having the same type of activity. In some aspects, pharmaceutical compositions exist in unsolvated form or in solvated forms with pharmaceutically acceptable solvents such as water, ethanol, and the like. The solvated forms of the pharmaceutical compositions are also considered to be disclosed herein.

[00173] In some aspects, a pharmaceutical composition exists as a tautomer. All tautomers are included within the scope of the agents presented herein. As such, it is to be understood that a pharmaceutical composition or a salt thereof may exhibit the phenomenon of tautomerism whereby two chemical compounds that are capable of facile interconversion by exchanging a hydrogen atom between two atoms, to either of which it forms a covalent bond. Since the tautomeric compounds exist in mobile equilibrium with each other they can be regarded as different isomeric forms of the same compound.

[00174] In some aspects, a pharmaceutical composition exists as an enantiomer, diastereomer, or other stereoisomeric form. The agents disclosed herein include all enantiomeric, diastereomeric, and epimeric forms as well as mixtures thereof.

[00175] In some aspects, pharmaceutical compositions described herein can be prepared as prodrugs. A "prodrug" refers to an agent that is converted into the parent drug *in vivo*. Prodrugs are often useful because, in some situations, they can be easier to administer than the parent drug. They may, for instance, be bioavailable by oral administration whereas the parent is not. The prodrug may also have improved solubility in pharmaceutical compositions over the parent drug. In certain embodiments, upon *in vivo* administration, a prodrug is chemically converted to the biologically, pharmaceutically or therapeutically active form of the pharmaceutical composition. In certain embodiments, a prodrug is enzymatically metabolized by one or more steps or processes to the biologically, pharmaceutically or therapeutically active form of the pharmaceutical composition.

Kits

[00176] Described herein, in some aspects, are kits for using the system, the composition, or the pharmaceutical composition described herein. In some aspects, the kits disclosed herein may be used to treat a disease or condition in a subject. In some aspects, the kit comprises an assemblage of materials or components apart from the system, the composition, or the pharmaceutical composition. In some aspects, the kit comprises the components for assaying and selecting for suitable guide nucleic acid or donor strand for treating a disease or a condition.

In some aspects, the kit comprises components for performing assays such as enzyme-linked immunosorbent assay (ELISA), single-molecular array (Simoa), PCR, or qPCR. The exact nature of the components configured in the kit depends on its intended purpose. For example, some embodiments are configured for the purpose of treating a disease or condition disclosed herein in a subject. In some aspects, the kit is configured particularly for the purpose of treating mammalian subjects. In some aspects, the kit is configured particularly for the purpose of treating human subjects.

[00177] Instructions for use may be included in the kit. In some aspects, the kit comprises instructions for administering the composition to a subject in need thereof. In some aspects, the kit comprises instructions for further engineering the system described herein. In some aspects, the kit comprises instructions for thawing or otherwise restoring biological activity of at least one component of the system, which may have been cryopreserved or lyophilized during storage or transportation. In some aspects, the kit comprises instructions for measuring efficacy for its intended purpose (e.g., therapeutic efficacy if used for treating a subject).

[00178] The kit may comprise a system or composition disclosed herein, and a container. The composition may be a pharmaceutical composition.

Methods

[00179] Described herein are methods such as methods of modifying a target nucleic acid. Described herein are methods such as methods of gene editing or gene replacement. The method may include use of any aspect of composition described herein such as an endonuclease, ligase, guide nucleic acid, integrating nucleic acid, system, kit, or pharmaceutical composition.

Gene editing or replacement

[00180] Disclosed herein are editing methods such as gene editing methods or nucleic acid editing methods. The editing tools and methods disclosed herein may be useful for genetic enhancement, genetic correction, treatment of a disease, development of research tools, or for disease diagnosis. The methods may be performed for therapeutic, agricultural, industrial, and research purposes. The editing method may include contacting a target nucleic acid with an editing system and a ligase. The target nucleic acid may be double-stranded. The target nucleic acid may include a host or cell genome. The target nucleic acid may include a pathogen genome in a host. The target nucleic acid may be part of a gene, or may include a non-gene or intergenic sequence. The target nucleic acid may reside in a nucleus of a cell. The target nucleic acid may include chromatin, euchromatin, or heterochromatin. The target nucleic acid may comprise DNA. The methods referred to herein as gene editing methods or genome editing methods may be useful for nucleic acid editing without necessarily being limited to editing of a certain gene.

The method may include replacing a target nucleic acid sequence with a sequence of an integrating nucleic acid. The method may be performed *in vitro*. The method may be performed *in vivo*. The method may be performed in a cell. The editing may be performed without homologous recombination. The editing may be performed without prior insertion into host genome.

[00181] Disclosed herein, in some aspects, are editing methods. The method may include editing a nucleic acid. The nucleic acid may be in a cell. The editing may be performed using a DNA ligase. The editing may be performed using a CRISPR protein. The editing may be performed using a CRISPR protein or DNA ligase without any significant chemical interaction with an endogenous factor. The editing may be performed using a CRISPR protein or DNA ligase without any significant chemical interaction with a polymerase such as a DNA polymerase. In some aspects, the editing may be performed using an endonuclease (e.g., a Cas endonuclease) described herein or DNA ligase, where the endonuclease and the DNA ligase are coupled. For example, the endonuclease and the DNA ligase can be covalently coupled as a fusion protein for performing the editing. The method may include editing a nucleic acid in a cell, wherein the editing is performed using a Cas endonuclease without any significant chemical interaction with an endogenous factor or polymerase. The method may include editing a nucleic acid in a cell, wherein the editing is performed using a Cas endonuclease without any significant chemical interaction with endogenous cellular components of NHEJ or HDR. The editing method may exclude polymerization or in-cell synthesis of a nucleic acid. For example, the method may exclude in-cell synthesis from a template on a guide nucleic acid.

[00182] The editing may be performed, in some aspects, solely by factors exogenous to the cell. The exogenous factors may be added to the cell or are encoded by a nucleic acid added to the cell. In some aspects, the exogenous factors are added to the cell. In some aspects, the exogenous factors encoded by a nucleic acid added to the cell. The factors may include a Cas endonuclease and a DNA ligase. The Cas endonuclease may be or include a DNA-binding protein.

[00183] The editing may include replacing a nucleotide or nucleotide sequence within a target nucleic acid. The editing may include replacing a nucleotide. The editing may include replacing a nucleotide sequence. The nucleotide or nucleotide sequence may be replaced with an integrating nucleic acid. The editing may include replacing a nucleotide or nucleotide sequence of the nucleic acid with an integrating nucleic acid. In some aspects, replacing the nucleotide comprises breaking a phosphodiester bond of the nucleic acid and forming a new phosphodiester bond with the integrating nucleic acid. In some aspects, the replacement is performed at a replacement site within the nucleic acid, without leaving a remaining nick or

strand break in the nucleic acid at the replacement site. In some aspects, the editing generates an edited nucleic acid comprising an edited region flanked by phosphodiester bonds to unedited regions of the edited nucleic acid.

[00184] Described herein, in some aspects, is a method for correcting a gene or modifying gene expression in a cell. In some aspects, the method comprises contacting the cell with a system or composition described herein. In some aspects, the method comprises delivering a heterologous polynucleotide into the cell, where the heterologous polynucleotide encodes at least one component of system. In some aspects, the system described herein can introduce a donor strand into a genomic locus. In some aspects, the system can introduce the donor strand without the need of endogenous machinery of the cell. In some aspects, the system can introduce the donor strand without the need to synchronize cell cycling. In some aspects, the system can introduce the donor strand in non-dividing cell or slow dividing cell. Such technical aspect can be especially useful for correcting genetic mutation in non-dividing cell or slow dividing cell for treating a disease or condition.

[00185] The method may include editing a nucleic acid of a cell. In some embodiments, the cell is quiescent or senescent cell. The cell may be quiescent. The cell may be senescent. In some aspects, the cell is not actively dividing. The cell may have a low dNTP concentration relative to other cells or cell types. Some examples of cells may include a neuron, myocyte, cardiomyocyte, or osteocyte. The cell may include a neuron. The cell may include a myocyte. The cell may include a cardiomyocyte. The cell may include an osteocyte. The cell may include an eye cell.

[00186] The cell may include a stem cell such as an embryonic stem cell, or such as an adult stem cell. The cell may be a circulating cell such as a blood cell. The cell may include a bone marrow cell. The cell may be an immune cell. The cell may be an innate immune cell.

[00187] The cell may be an airway cell. The cell may be a lung cell. The cell may be a bronchial cell. The cell may be an endothelial cell.

[00188] Described herein, in some aspects, is an editing method, comprising: editing a nucleic acid in a cell, wherein the editing is performed using a CRISPR protein (e.g. an RNA-guided endonuclease such as a Cas endonuclease) without any significant chemical interaction with an endogenous factor or polymerase. In some embodiments, the editing is performed solely by factors exogenous to the cell. In some embodiments, the exogenous factors are added to the cell or are encoded by a nucleic acid added to the cell.

[00189] In some embodiments, the editing is performed using a DNA ligase. In some embodiments, the editing comprises replacing a nucleotide or nucleotide sequence of the nucleic acid with an integrating nucleic acid. In some embodiments, replacing the nucleotide

comprises breaking a phosphodiester bond of the nucleic acid and forming a new phosphodiester bond with the integrating nucleic acid. In some embodiments, the replacement is performed at a replacement site within the nucleic acid, without leaving a nick or strand break in the nucleic acid at the replacement site. In some embodiments, the editing generates an edited nucleic acid comprising an edited region flanked by phosphodiester bonds to unedited regions of the edited nucleic acid.

[00190] Some aspects include a method for modifying a cell comprising contacting a cell with a system or composition such as a pharmaceutical composition disclosed herein. In some aspects, the cell is not a dividing cell. The integrating nucleic acid may be inserted into the genomic locus of the cell independent of endogenous non-homologous end joining (NHEJ) and independent of endogenous homology-directed repair (HDR).

[00191] In some aspects, described herein is a method for modifying or replacing a nucleotide or nucleotide sequence in a cell by contacting the cell with the system or composition described herein, where the system or composition comprises a guide nucleic acid comprising: a spacer complementary to a region of a genomic locus of a genomic strand; a scaffold for complexing with an endonuclease; an optional donor binding site that is at least partially complementary to an integrating nucleic acid; and a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus. In some embodiments, the guide nucleic acid comprises the donor binding site is complexed with the integrating nucleic acid. The complexing between the guide nucleic acid and the integrating nucleic acid can occur *in vivo* or *in vitro*. In some embodiments, the flap binding site can be complexed with a genomic flap generated by the endonuclease cleaving the genomic strand. The complexing between the flap binding site and the genomic flap can bring the integrating nucleic acid to close proximity to the cleaved genomic strand. The decreased proximity between the donor nucleic and the cleaved genomic strand can increase editing efficiency, decrease off-target effect, or decrease introduction of unwanted mutations such as indels. In such case, the integrating nucleic acid can replace one strand of the cleaved genomic strand, thus editing or correcting the cleaved genomic strand. **Fig. 1A-Fig. 1C** illustrate the complexing between the guide nucleic acid and the integrating nucleic acid described herein, where the complexing between the guide nucleic acid and the integrating nucleic acid brings the integrating nucleic acid to close proximity to the cleaved genomic strand. In some embodiments, the integrating nucleic acid comprises a 5' end to be ligated to a 3' terminus of a genomic strand generated by an endonuclease cleaving the genomic strand. In some embodiments, the integrating nucleic acid comprises a 3' end to be ligated to a 5' terminus of a genomic strand generated by an endonuclease cleaving the genomic strand. In some embodiments, the endonuclease can be a

fusion protein described herein. For example, the endonuclease can be fused to a DNA ligase described herein, where the endonuclease and DNA ligase fusion can cleave the genomic strand and ligate the integrating nucleic acid to the cleaved genomic strand with increased efficiency.

[00192] In some embodiments, the integrating nucleic acid is double stranded or partially double stranded, where the integrating nucleic acid can replace both strands of the cleaved genomic strand. In such case, the integrating nucleic acid can comprise single stranded guide binding site to be complexed with a guide nucleic acid comprising the donor binding site. The guide binding site can locate at 5' end of the integrating nucleic acid. The guide binding site can locate at 3' end of the integrating nucleic acid. The guide binding site can locate at both 5' end and 3' end of the integrating nucleic acid. **Fig. 2A-Fig. 2C** illustrate a double stranded integrating nucleic acid comprising the guide binding site at both 5' end and 3' end of the integrating nucleic acid, where the integrating nucleic acid can edit and replace the cleaved genomic strand.

[00193] In some embodiments, the integrating nucleic acid is double stranded or partially double stranded, where the integrating nucleic acid comprises a flap binding site and a guide binding site. In such case, the guide binding site can complex with the donor binding site of the guide nucleic acid. **Fig. 3A** illustrates such arrangement, where the integrating nucleic acid (and not the guide nucleic acid) can be complexed with the genomic flap to bring the integrating nucleic acid to close proximity to the cleaved genomic strand. In some embodiments, the donor nucleic acid comprises two flap binding sites to be complexed with two different genomic flaps. **Fig. 4A** illustrates such arrangement, where the integrating nucleic acid (and not the guide nucleic acid) can be complexed with the two genomic flaps to bring the integrating nucleic acid to close proximity to the two cleaved genomic strand.

[00194] In some embodiments, the integrating nucleic acid comprises the guide binding site, where the guide binding site can be complexed with the donor binding site of the guide nucleic acid. The guide nucleic acid can comprise the flap binding site to be complexed with the genomic flap at the cleaved genomic strand. As shown in **Fig. 5A**, the guide nucleic acid brings the integrating nucleic acid to close proximity to the cleaved genomic strand for editing and replacing the cleaved genomic strand with the integrating nucleic acid. In some embodiments, the integrating nucleic acid can be double strand and comprises the two guide binding sites to be complexed with two different guide nucleic acids. **Fig. 6A** illustrates such arrangement, where the two guide nucleic acids bring the integrating nucleic acid to close proximity to two cleaved genomic strands.

[00195] In some aspects, described herein is a method for modifying or replacing a nucleotide or nucleotide sequence in a cell by contacting the cell with the system or composition described

herein, where the system or composition comprises a guide nucleic acid comprising a spacer complementary to a region of a genomic locus of a genomic strand; a scaffold for complexing with an endonuclease, and an optional donor binding site that is at least partially complementary to a splinting nucleic acid. In some embodiments, the system or composition comprises an integrating nucleic acid, where the integrating nucleic acid can be ligated into the cleaved or nicked genomic strand. In some embodiments, the integrating nucleic acid comprises a 5' end to be ligated to a 3' terminus of the genomic strand generated by an endonuclease. In some embodiments, the integrating nucleic acid comprises a 3' end to be ligated to a 5' terminus of the genomic strand generated by an endonuclease. In some embodiments, the system or composition comprises a splinting nucleic acid comprising a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus, and comprising an optional guide binding site that is at least partially complementary to a guide nucleic acid. In some embodiments, the splinting nucleic acid may include a guide binding site. The guide binding site may be complementary to a guide nucleic acid. The guide binding site may be partially complementary to a guide nucleic acid. The splinting nucleic acid may include a donor binding site. The donor binding site may be complementary to an integrating nucleic acid. The donor binding site may be partially complementary to an integrating nucleic acid. The splinting strand may be or include DNA. The splinting strand may be or include RNA. The splinting nucleic acid may be included as part of an integrating nucleic acid. The splinting nucleic acid may be included as a strand of a double stranded integrating nucleic acid.

[00196] In some embodiments, the method described herein decreases proximity between the integrating nucleic acid and the cleaved or nicked site. In some embodiments, the decreased proximity between the integrating nucleic acid and the cleaved or nicked site increases gene editing rate by at least 0.1 fold, 0.2 fold, 0.5 fold, 1.0 fold, 2.0 fold, 5.0 fold, 10.0 fold, or more compared to a gene editing rate without using a composition or a replacer described herein. In some embodiments, the decreased proximity between the integrating nucleic acid and the cleaved or nicked site decreases introduction of unwanted mutation such as indel by at least 0.1 fold, 0.2 fold, 0.5 fold, 1.0 fold, 2.0 fold, 5.0 fold, 10.0 fold, or more compared to a introduction of unwanted mutation without using a composition or a replacer described herein. In some embodiments, the decreased proximity between the integrating nucleic acid and the cleaved or nicked site decreases off-target editing by at least 0.1 fold, 0.2 fold, 0.5 fold, 1.0 fold, 2.0 fold, 5.0 fold, 10.0 fold, or more compared to off-target editing without using a composition or a replacer described herein.

[00197] In some aspects, the method edits a gene. In some aspects, the method replaces a gene. In some aspects, the method removes a gene. In some aspects, the method introduces a methylated nucleotide into the target nucleic acid. In some aspects, the method introduces an unmethylated nucleotide into the target nucleic acid.

[00198] The method may be used to edit a nucleic acid in a plant cell. Some aspects include enhancing a plant. Some examples of plant enhancement may include editing of a disease susceptibility gene or introducing an herbicide resistance gene. An example of a disease susceptibility gene may include bacterial leaf streak disease susceptibility gene OsSULTR3;6 in rice. An example of introducing an herbicide resistance gene may include editing of acetolactate synthase in potato for herbicide resistance

Treatment

[00199] A method such as a gene editing method may be useful for treatment of a disease or disorder. The disease or disorder may be genetic. The treatment may be of a diseased or damaged cell. The disease may include a genetic disease, cancer, or an infection. The treatment may include administration of a composition disclosed herein to a subject in need thereof. The subject in need may include a subject identified as having a disease or disorder.

[00200] The methods described herein may be useful for treating a genetic disease. The genetic disease may be caused by a DNA mutation such as a point mutation, a deletion, an insertion, a duplication, or a repeat, relative to normal non-diseased DNA. The treatment may correct the mutation. Some examples of genetic diseases may include Angelman syndrome, Canavan disease, Charcot–Marie–Tooth disease, color blindness, cri du chat syndrome, cystic fibrosis, DiGeorge syndrome, Duchenne muscular dystrophy, familial hypercholesterolemia, haemochromatosis type 1, hemophilia, neurofibromatosis, phenylketonuria, polycystic kidney disease, Prader–Willi syndrome, sickle cell disease, spinal muscular atrophy, or Tay–Sachs disease. Some examples of diseases that may be treated using a method herein may include sickle cell disease, beta thalassemia, familial hypercholesterolemia (e.g. PCSK9 disruption), alpha I antitrypsin deficiency, phenylketonuria, cystic fibrosis, tyrosinemia, arginase I deficiency, Wilson’s disease, a repeat expansion disorder, hemophilia (e.g. insertion of Factor IX at ALB in a hepatocyte), Duchenne muscular dystrophy. Some examples of repeat expansion disorders like Huntington's disease, Amyotrophic lateral sclerosis/frontotemporal dementia, Friedreich ataxia, Fragile X Syndrome. The method may be included in immuno-oncology, such as for T-cell engineering or in cancer treatment.

[00201] Two non-limiting examples of genetic diseases for which efficient and precise editing of slowly dividing and nondividing cells is beneficial for therapeutic gene therapy are sickle cell anemia (SCA) and alpha-1 antitrypsin deficiency (AATD). Sickle cell anemia is caused by

the E6V missense mutation in the HBB gene resulting in aggregation of mutant beta-globin protein and ‘sickling’ of red blood cells. Autologous gene therapies using hematopoietic stem cells with corrected HBB alleles have been proposed as curative treatments for SCA. While expansion of *ex vivo* HSC cultures can be induced using cytokine cocktails, HSCs in the human body typically reside in niches within the bone marrow where they exist in a quiescent or slowly dividing state. AATD is most commonly caused by the E366K missense mutation in the SERPINA1 gene which encodes alpha-1 antitrypsin, a serine protease inhibitor secreted by hepatocytes. Mutant AAT is misfolded, forming aggregates in the endoplasmic reticulum of the hepatocytes rather than being secreted, ultimately leading to liver disease. Although hepatocytes possess the ability to rapidly proliferate in response to liver damage, their life cycles are typically spent in a state of quiescence. As such, high efficiency *in vivo* editing of these two disorders necessitates a novel gene therapy platform which can effectively perform precise edits in nondividing or slowly dividing cells.

[00202] Some aspects include a method for treating a disease or condition in subject in need thereof comprising: (a) contacting a cell of the subject with a system or composition such as a pharmaceutical composition disclosed herein; and (b) replacing a genomic locus in a cell with an integrating nucleic acid, thereby treating the disease or condition in the subject. In some aspects, the cell is not a dividing cell. In some aspects, the integrating nucleic acid is inserted into the genomic locus of the cell independent of endogenous non-homologous end joining (NHEJ) and independent of endogenous homology-directed repair (HDR).

[00203] In some embodiments, the method described herein decreases proximity between the integrating nucleic acid and the cleaved or nicked site, where the decreased proximity between the integrating nucleic acid and the cleaved or nicked site increases gene editing rate by at least 0.1 fold, 0.2 fold, 0.5 fold, 1.0 fold, 2.0 fold, 5.0 fold, 10.0 fold, or more compared to a gene editing rate without using a composition or a replacer described herein. In some embodiments, the decreased proximity between the integrating nucleic acid and the cleaved or nicked site increases therapeutic efficacy (e.g., by increasing gene editing rate) by at least 0.1 fold, 0.2 fold, 0.5 fold, 1.0 fold, 2.0 fold, 5.0 fold, 10.0 fold, or more compared to a therapeutic efficacy without using a composition or a replacer described herein.

Delivery

[00204] Described herein, in some aspects, are methods of delivering the system described herein to a cell. In some aspects, the method comprises delivering directly or indirectly at least one component of the system to the cell. In some aspects, the method comprises delivering the cell with at least one heterologous polynucleotide, where the cell can then express the at least one component of the system. In some aspects, the at least one heterologous polynucleotide can

be delivered into the cell via any of the transfection methods described herein. In some aspects, the at least one heterologous polynucleotide can be delivered into the cell via the use of expression vectors such as viral vectors. In the context of an expression vector, the vector can be readily introduced into the cell described herein by any method in the art. For example, the expression vector can be transferred into the cell by physical, chemical, or biological means.

[00205] Physical methods for introducing the oligonucleotide or vector encoding the oligonucleotide into the cell can include calcium phosphate precipitation, lipofection, particle bombardment, microinjection, gene gun, electroporation, and the like. Methods for producing cells comprising vectors and/or exogenous nucleic acids are suitable for methods herein. One method for the introduction of oligonucleotide or vector encoding the oligonucleotide into a host cell is calcium phosphate transfection.

[00206] Chemical means for introducing the oligonucleotide or vector encoding the oligonucleotide into the cell can include colloidal dispersion systems, such as macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, spherical nucleic acid (SNA), liposomes, or lipid nanoparticles. An exemplary colloidal system for use as a delivery vehicle *in vitro* and *in vivo* is a liposome (e.g., an artificial membrane vesicle). Other methods of state-of-the-art targeted delivery of nucleic acids are available, such as delivery of oligonucleotide or vector encoding the oligonucleotide with targeted nanoparticles or other suitable sub-micron sized delivery system.

[00207] In the case where a non-viral delivery system is utilized, an exemplary delivery vehicle is a liposome. The use of lipid formulations is contemplated for the introduction of the oligonucleotide or vector encoding the oligonucleotide into a cell (*in vitro*, *ex vivo* or *in vivo*). In another aspect, the oligonucleotide or vector encoding the oligonucleotide can be associated with a lipid. The oligonucleotide or vector encoding the oligonucleotide associated with a lipid, In some aspects, is encapsulated in the aqueous interior of a liposome, interspersed within the lipid bilayer of a liposome, attached to a liposome via a linking molecule that is associated with both the liposome and the oligonucleotide, entrapped in a liposome, complexed with a liposome, dispersed in a solution containing a lipid, mixed with a lipid, combined with a lipid, contained as a suspension in a lipid, contained or complexed with a micelle, or otherwise associated with a lipid. Lipid, lipid/DNA or lipid/expression vector associated compositions are not limited to any particular structure in solution. For example, In some aspects, they are present in a bilayer structure, as micelles, or with a “collapsed” structure. Alternately, they may be simply interspersed in a solution, possibly forming aggregates that are not uniform in size or shape. Lipids are fatty substances which are, In some aspects, naturally occurring or synthetic

lipids. For example, lipids include the fatty droplets that naturally occur in the cytoplasm as well as the class of compounds which contain long-chain aliphatic hydrocarbons and their derivatives, such as fatty acids, alcohols, amines, amino alcohols, and aldehydes.

[00208] Lipids suitable for use are obtained from commercial sources. Stock solutions of lipids in chloroform or chloroform/methanol are often stored at about -20 °C. Chloroform is used as the only solvent since it is more readily evaporated than methanol. "Liposome" is a generic term encompassing a variety of single and multilamellar lipid vehicles formed by the generation of enclosed lipid bilayers or aggregates. Liposomes are often characterized as having vesicular structures with a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers. However, compositions that have different structures in solution than the normal vesicular structure are also encompassed. For example, the lipids, in some aspects, assume a micellar structure or merely exist as nonuniform aggregates of lipid molecules. Also contemplated are lipofectamine-nucleic acid complexes.

[00209] In some cases, non-viral delivery method comprises lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, exosomes, polycation or lipid:cargo conjugates (or aggregates), naked polypeptide (e.g., recombinant polypeptides), naked DNA, artificial virions, and agent-enhanced uptake of polypeptide or DNA. In some aspects, the delivery method comprises conjugating or encapsulating the compositions or the oligonucleotides described herein with at least one polymer such as natural polymer or synthetic materials. The polymer can be biocompatible or biodegradable. Non-limiting examples of suitable biocompatible, biodegradable synthetic polymers can include aliphatic polyesters, poly(amino acids), copoly(ether-esters), polyalkylenes oxalates, polyamides, poly(iminocarbonates), polyorthoesters, polyoxaesters, polyamidoesters, polyoxaesters containing amine groups, and poly(anhydrides). Such synthetic polymers can be homopolymers or copolymers (e.g., random, block, segmented, graft) of a plurality of different monomers, e.g., two or more of lactic acid, lactide, glycolic acid, glycolide, epsilon-caprolactone, trimethylene carbonate, p-dioxanone, etc. In an example, the scaffold can be comprised of a polymer comprising glycolic acid and lactic acid, such as those with a ratio of glycolic acid to lactic acid of 90/10 or 5/95. Non-limiting examples of naturally occurring biocompatible, biodegradable polymers can include glycoproteins, proteoglycans, polysaccharides, glycosamineoglycan (GAG) and fragment(s) derived from these components, elastin, laminins, decorin,

fibrinogen/fibrin, fibronectins, osteopontin, tenascins, hyaluronic acid, collagen, chondroitin sulfate, heparin, heparan sulfate, ORC, carboxymethyl cellulose, and chitin.

[00210] In some cases, the oligonucleotide or vector encoding the oligonucleotide described herein can be packaged and delivered to the cell via extracellular vesicles. The extracellular vesicles can be any membrane-bound particles. In some aspects, the extracellular vesicles can be any membrane-bound particles secreted by at least one cell. In some instances, the extracellular vesicles can be any membrane-bound particles synthesized *in vitro*. In some instances, the extracellular vesicles can be any membrane-bound particles synthesized without a cell. In some cases, the extracellular vesicles can be exosomes, microvesicles, retrovirus-like particles, apoptotic bodies, apoptosomes, oncosomes, exophers, enveloped viruses, exomeres, or other very large extracellular vesicles.

[00211] In aspects, the system described herein or the at least one heterologous polynucleotide encoding the system described herein can be delivered into a cell as a vector such as a viral vector. Viral vectors, and especially retroviral vectors, have become the most widely used method for inserting genes into mammalian, e.g., human cells. Other viral vectors, in some embodiments, are derived from lentivirus, poxviruses, herpes simplex virus I, adenoviruses and adeno-associated viruses, and the like. Exemplary viral vectors include retroviral vectors, adenoviral vectors, adeno-associated viral vectors (AAVs), pox vectors, parvoviral vectors, baculovirus vectors, measles viral vectors, or herpes simplex virus vectors (HSVs). In some instances, the retroviral vectors include gamma-retroviral vectors such as vectors derived from the Moloney Murine Leukemia Virus (MoMLV, MMLV, MuLV, or MLV) or the Murine Stem cell Virus (MSCV) genome. In some instances, the retroviral vectors also include lentiviral vectors such as those derived from the human immunodeficiency virus (HIV) genome. In some instances, AAV vectors include AAV1, AAV2, AAV4, AAV5, AAV6, AAV7, AAV8, or AAV9 serotype. In some instances, viral vector is a chimeric viral vector, comprising viral portions from two or more viruses. In additional instances, the viral vector is a recombinant viral vector.

[00212] In some cases, the at least one heterologous polynucleotide encoding the system described herein can be administered to the subject in need thereof via the use of the transgenic cells generated by introduction of the at least one heterologous polynucleotide first into allogeneic or autologous cells. In some cases, the cell can be isolated. In some aspects, the cell can be isolated from the subject.

Subjects and cells

[00213] The methods described herein may involve cells. For example, a composition may be delivered to a cell to edit a nucleic acid in the cell. The aspects delivered to the cell may be

heterologous to the cell. "Heterologous" may include anything that does not exist in the cell in its natural state.

[00214] Any cell or cell type may be used. Examples of cells or cell types may include stem cells, red blood cells, white blood cells, platelets, nerve cells, neuroglial cells, muscle cells, cartilage cells, bone cells, skin cells, endothelial cells, epithelial cells, fat cells, or sex cells. The cell may include a stem cell. The cell may include a bone cell. The cell may include a blood cell. The cell may include a sperm cell. The cell may include an egg cell. The cell may include a fat cell. The cell may include a nerve cell. The cell may include a muscle cell. The cell may include an endocrine cell. The cell may include an endothelial cell. The cell may include a pancreatic cell.

[00215] The cell may be eukaryotic. The cell may be a plant cell. The cell may be an animal cell. The cell may be protozoan. The cell may be a fungal cell. The cell may be prokaryotic. The cell may be a bacterial cell. The cell may be an archaeon cell. The cell may be from a cell line. The cell may be part of a subject. The cell may be separated from a subject. The cell may be an autologous cell of a subject. The cell may be an allogenic cell of a subject.

[00216] The cell may include a diseased cell. The cell may include a cancer cell. The cell may be infected. The cell may be damaged. The cell may be a pathogen such as a fungal pathogen.

[00217] The methods described herein may involve a subject. For example, a composition may be delivered to the subject. Some aspects of the methods described herein include treatment of the subject. Non-limiting examples of subjects include vertebrates, animals, mammals, dogs, cats, cattle, rodents, mice, rats, primates, monkeys, and humans. The subject may be an invertebrate. The subject may be an arthropod. The subject may be a vertebrate. The subject may be an animal. The subject may be a fish. The subject may be a reptile. The subject may be a mammal. The subject may be a dog. The subject may be a cat. The subject may be a cattle. The subject may be a rodent. The subject may be a mouse. The subject may be a rat. The subject may be a primate. The subject may be a non-human primate. The subject may be a monkey. The subject may be an animal, a mammal, a dog, a cat, cattle, a rodent, a mouse, a rat, a primate, or a monkey. The subject may be a human.

[00218] The subject may be a non-animal subject. For example, the subject may include a plant. Examples of plants may include trees, flowers, shrubs, or grasses. The subject may include a crop. Examples of crops may include almond, apricot, apple, artichoke, banana, barley, beet, blackberry, blueberry, broccoli, Brussels sprout, cabbage, cannabis, capsicum, carrot, celery, chard, cherry, citrus, corn, cucurbit, date, fig, garlic, grape, herb, spice, kale, lettuce, oil palm, olive, onion, pea, pear, peach, peanut, papaya, parsnip, pecan, persimmon, plum, pomegranate,

potato, quince, radish, raspberry, rose, rice, sloe, sorghum, soybean, spinach, strawberry, sweet potato, tobacco, tomato, turnip greens, walnut, or wheat.

Definitions

[00219] Use of absolute or sequential terms, for example, “will,” “will not,” “shall,” “shall not,” “must,” “must not,” “first,” “initially,” “next,” “subsequently,” “before,” “after,” “lastly,” and “finally,” are not meant to limit scope of the present embodiments disclosed herein but as exemplary.

[00220] As used herein, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. Furthermore, to the extent that the terms “including,” “includes,” “having,” “has,” “with,” or variants thereof are used in either the detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term “comprising.”

[00221] As used herein, the phrases “at least one,” “one or more,” and “and/or” are open-ended expressions that are both conjunctive and disjunctive in operation. For example, each of the expressions “at least one of A, B and C”, “at least one of A, B, or C”, “one or more of A, B, and C”, “one or more of A, B, or C” and “A, B, and/or C” means A alone, B alone, C alone, A and B together, A and C together, B and C together, or A, B and C together.

[00222] As used herein, “or” may refer to “and”, “or,” or “and/or” and may be used both exclusively and inclusively. For example, the term “A or B” may refer to “A or B”, “A but not B”, “B but not A”, and “A and B”. In some cases, context may dictate a particular meaning.

[00223] Any systems, methods, software, and platforms described herein are modular. Accordingly, terms such as “first” and “second” do not necessarily imply priority, order of importance, or order of acts.

[00224] The term “about” when referring to a number or a numerical range means that the number or numerical range referred to is an approximation within experimental variability (or within statistical experimental error), and the number or numerical range may vary from, for example, from 1% to 15% of the stated number or numerical range. In examples, the term “about” refers to $\pm 10\%$ of a stated number or value.

[00225] The terms “increased”, “increasing”, or “increase” are used herein to generally mean an increase by a statically significant amount. In some aspects, the terms “increased,” or “increase,” mean an increase of at least 10% as compared to a reference level, for example an increase of at least about 10%, at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and including a 100% increase or any increase between 10-100% as compared to a reference level, standard, or control. Other examples of “increase” include an

increase of at least 2-fold, at least 5-fold, at least 10-fold, at least 20-fold, at least 50-fold, at least 100-fold, at least 1000-fold or more as compared to a reference level.

[00226] The terms “decreased”, “decreasing”, or “decrease” are used herein generally to mean a decrease by a statistically significant amount. In some aspects, “decreased” or “decrease” means a reduction by at least 10% as compared to a reference level, for example a decrease by at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90% or up to and including a 100% decrease (e.g., absent level or non-detectable level as compared to a reference level), or any decrease between 10-100% as compared to a reference level. In the context of a marker or symptom, by these terms is meant a statistically significant decrease in such level. The decrease can be, for example, at least 10%, at least 20%, at least 30%, at least 40% or more, and is preferably down to a level accepted as within the range of normal for an individual without a given disease.

[00227] Where sequences are provided, nucleic acids containing phosphorothioate bonds between nucleotides are signified with an asterisk (*). 2'-O-methyl nucleotides are signified with a lowercase “m” in front of the nucleotide, for example mC instead of C. The code “/5Phos/” in front of a nucleotide sequence indicates that the sequence is phosphorylated at the 5' end. Locked nucleic acid (LNA) nucleotides comprising a methylene bridge connecting the 2' oxygen and 4' carbon are signified with a “+” in front of the nucleotide, for example +C instead of C.

EMBODIMENTS

[00228] Some aspects include an embodiment, or a component thereof, provided below:

[00229] Embodiment 1. Described herein, in some aspects, is a composition, comprising: a DNA-binding protein coupled to a DNA ligase.

[00230] Embodiment 2. The composition of Embodiment 1, wherein the coupling is covalent.

[00231] Embodiment 3. The composition of Embodiment 2, comprising a fusion protein comprising the DNA-binding protein and the DNA ligase.

[00232] Embodiment 4. The composition of Embodiment 3, wherein the DNA-binding protein is amino (N)-terminal relative to the DNA ligase within the fusion protein.

[00233] Embodiment 5. The composition of Embodiment 3, wherein the DNA-binding protein is carboxy (C)-terminal relative to the DNA ligase within the fusion protein.

[00234] Embodiment 6. The composition of any one of Embodiments 2-5, wherein the connection comprises a linker comprising 1-100 amino acids.

[00235] Embodiment 7. The composition of Embodiment 1, wherein the coupling is non-covalent.

[00236] Embodiment 8. The composition of Embodiment 7, wherein the composition comprises a first polypeptide comprising at least part of the DNA-binding protein, and a second polypeptide comprising at least part of the DNA ligase, wherein the first and second polypeptides are non-covalently coupled.

[00237] Embodiment 9. The composition of Embodiment 8, wherein the first polypeptide comprises a first heterodimerization domain that binds a second heterodimerization domain, and wherein the second polypeptide comprises the second heterodimerization domain.

[00238] Embodiment 10. The composition of Embodiment 9, wherein the heterodimer domains comprise a leucine zipper, PDZ domain, streptavidin, streptavidin binding protein, foldon domain, hydrophobic moiety, or a functional binding fragment thereof.

[00239] Embodiment 11. The composition of Embodiment 8, wherein the first polypeptide comprises a first intein that binds a second intein, and wherein the second polypeptide comprises the second intein.

[00240] Embodiment 12. The composition of Embodiment 1, wherein the ligase comprises a hairpin binding motif, and wherein the DNA-binding protein and the DNA ligase are coupled with a nucleic acid comprising a scaffold that binds to the DNA-binding protein and a hairpin that binds to the hairpin binding motif.

[00241] Embodiment 13. The composition of Embodiment 12, wherein the hairpin binding motif comprises an MS2 coat protein (MCP) peptide, and wherein the hairpin comprises an MS2 hairpin.

[00242] Embodiment 14. The composition of Embodiment 1, wherein the DNA-binding protein and the DNA ligase are coupled with a heterobifunctional molecule comprising an endonuclease binding domain and a DNA ligase binding domain.

[00243] Embodiment 15. The composition of Embodiment 14, wherein the heterobifunctional molecule comprises a small molecule.

[00244] Embodiment 16. Described herein, in some aspects, is a composition comprising a cell containing a DNA-binding protein and a DNA ligase, both of which are heterologous to the cell.

[00245] Embodiment 17. The composition of any one of Embodiments 1-16, wherein the DNA-binding protein comprises a class II CRISPR/Cas endonuclease.

[00246] Embodiment 18. The composition of any one of Embodiments 1-17, wherein the DNA-binding protein comprises a Cas9 endonuclease.

[00247] Embodiment 19. The composition of any one of Embodiments 1-18, wherein the DNA-binding protein comprises a nickase.

[00248] Embodiment 20. The composition of any one of Embodiments 1-19, wherein the DNA-binding protein comprises an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 1-13, or a functional fragment thereof.

[00249] Embodiment 21. The composition of any one of Embodiments 1-20, wherein the DNA ligase ligates DNA strands base paired to a DNA splint.

[00250] Embodiment 22. The composition of any one of Embodiments 1-20, wherein the DNA ligase ligates DNA strands base paired to an RNA splint.

[00251] Embodiment 23. The composition of any one of Embodiments 1-22, wherein the DNA ligase comprises an amino acid sequence at least 80% identical to the amino acid sequence of any one of SEQ ID NOS: 55-96, or a functional fragment thereof.

[00252] Embodiment 24. The composition of any one of Embodiments 1-23, wherein the DNA-binding protein or the DNA ligase comprises a nuclear localization signal, chromatin modifying domain, cell penetrating peptide, or tag polypeptide.

[00253] Embodiment 25. The composition of any one of Embodiments 1-24, further comprising a guide RNA and an integrating nucleic acid.

[00254] Embodiment 26. One or more nucleic acids encoding the composition of any one of Embodiments 1-25.

[00255] Embodiment 27. A cell comprising the composition of any one of Embodiments 1-25, or comprising the one or more nucleic acids of Embodiment 26.

[00256] Embodiment 28. A system of nucleic acids comprising:

- a. a guide nucleic acid comprising:
 - i. a spacer complementary to a region of a genomic locus of a genomic strand,
 - ii. a scaffold for complexing with a DNA-binding protein,
 - iii. an optional donor binding site that is at least partially complementary to an integrating nucleic acid, and
 - iv. a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus; and
- b. an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by a DNA-binding protein.

[00257] Embodiment 29. A system of nucleic acids comprising:

- a. a guide nucleic acid comprising:
 - i. a spacer complementary to a region of a genomic locus of a genomic strand,
 - ii. a scaffold for complexing with a DNA-binding protein, and

- iii. an optional donor binding site that is at least partially complementary to a splinting nucleic acid;
- b. an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by a DNA-binding protein; and
- c. a splinting nucleic acid comprising a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus, and comprising an optional guide binding site that is at least partially complementary to a guide nucleic acid.

[00258] Embodiment 30. The system of Embodiment 28 or 29, wherein the genomic strand is in a cell.

[00259] Embodiment 31. The system of any one of Embodiments 28-30, wherein the splinting nucleic acid further comprises a donor binding site that is at least partially identical or complementary to a portion of the integrating nucleic acid.

[00260] Embodiment 32. The system of any one of Embodiment 28-31, wherein the guide nucleic acid comprises a sequence of linking nucleic acids between the scaffold and the donor binding site.

[00261] Embodiment 33. The system of any one of Embodiment 28-32, wherein the guide nucleic acid, the integrating nucleic acid, or the splinting nucleic acid comprises a modified internucleoside linkage.

[00262] Embodiment 34. The system of Embodiment 33, wherein the modified internucleoside linkage comprises a phosphorothioate linkage.

[00263] Embodiment 35. The system of Embodiment 33 or 34, wherein the modified internucleoside linkage is between any of the 4 terminal nucleosides at a 5' end or at a 3' end of the guide nucleic acid or the integrating nucleic acid.

[00264] Embodiment 36. The system of any one of Embodiments 28-35, wherein the guide nucleic acid, the integrating nucleic acid, or the splinting nucleic acid comprises a modified nucleoside.

[00265] Embodiment 37. The system of Embodiment 36, wherein the modified nucleoside comprises a locked nucleic acid (LNA), a 2' fluoro, a 2' O-alkyl, or a combination thereof.

[00266] Embodiment 38. The system of Embodiment 36 or 37, wherein the modified nucleoside is any of the 3 terminal nucleosides at a 5' end or at a 3' end of the guide nucleic acid or the integrating nucleic acid.

EXAMPLES

Example 1. Editing to convert BFP to GFP by Replacer 1

[00267] Components used to edit the blue fluorescent protein (BFP) gene stably integrated into HEK293 cells are co-delivered by lipid nanoparticle (LNP) transfection. The components include chemically synthesized guide RNAs (gRNAs), single-stranded DNA donors, and mRNA encoding protein effectors for Replacer 1 editing including nicking Cas9 (nCas9), a SplintR ligase and nuclear localization sequences (NLS). The gRNAs are synthesized by Agilent, the DNA donors are synthesized by IDT, and the mRNA is synthesized by TriLink or RiboPro. The gRNA, DNA donor, and mRNA are mixed and formulated into lipid nanoparticles prior to delivery to adherent cells in 96well plates. After 48 hours, the cells are detached from the plate by trypsinization and green fluorescent protein (GFP) fluorescence is measured using an Attune NxT flow cytometer to assess the percentage of BFP-to-GFP editing. Following the Replacer 1 editing format, the gRNAs contain a spacer, scaffold, donor binding site (DBS), and flap binding site (FBS). The gRNAs are delivered individually (1-sided Replacer 1) or as pairs with spacers targeting opposite strands of the genomic locus (2-sided Replacer 1). Some of the DBSs contain a mutation in the spacer region or in the protospacer adjacent motif region (SpPAMmut). The gRNAs contain 2'-O-methyl 3'-phosphorothioate nucleotides at the first three and last three positions. The DNA donors are delivered individually (1-sided Replacer 1) or in pairs (2-sided Replacer 1). Some donors have mutations in the spacer or protospacer adjacent motif (PAM) regions (SpPAMmut). Some donors have phosphorothioate bonds at the first three and last three positions. Some donors are recoded with silent mutations that change the nucleotide sequence but retain the amino acid sequence. The DNA donors are phosphorylated on the 5' end. In some conditions, the gRNAs and donor DNAs are annealed by a thermal cycler annealing reaction prior to LNP formulation. Plasmids can be used in the place of mRNA. **Table 12** details this experiment. Sequences corresponding to the names in the table may be found herein.

Table 12

Condi tion	Forward Guide	Reverse Guide	Top Donor	Bottom Donor	Ligase	Anneal both sides?
1	Rep1.BFP.Fwd Guide		Rep1.BFP2GFP .TopDonor.5P		NLS- nCas9- linker- SplintR- bpNLS	N/A
2		Rep1.BFP .RevGuide		Rep1.BFP2G FP.BotDonor. 5P	NLS- nCas9- linker- SplintR- bpNLS	N/A

Condition	Forward Guide	Reverse Guide	Top Donor	Bottom Donor	Ligase	Anneal both sides?
3	Rep1.BFP.Fwd Guide	Rep1.BFP .RevGuide	Rep1.BFP2GFP .TopDonor.5P	Rep1.BFP2GFP.BotDonor.5P	NLS-nCas9-linker-SplintR-bpNLS	Yes
4	Rep1.BFP.Fwd Guide	Rep1.BFP .RevGuide	Rep1.BFP2GFP .TopDonor.5P	Rep1.BFP2GFP.BotDonor.5P	NLS-nCas9-linker-SplintR-bpNLS	No
5	Rep1.BFP.Fwd Guide	Rep1.BFP .RevGuide	Rep1.BFP2GFP .TopDonor.Recoded.5P	Rep1.BFP2GFP.BotDonor.Recoded.5P	NLS-nCas9-linker-SplintR-bpNLS	Yes
6	Rep1.BFP.Fwd Guide.SpPAMmut	Rep1.BFP .RevGuide .SpPAMmut	Rep1.BFP2GFP .TopDonor.SpPAMmut.Recoded.5P	Rep1.BFP2GFP.BotDonor.SpPAMmut.Recoded.5P	NLS-nCas9-linker-SplintR-bpNLS	Yes
7	Rep1.BFP.Fwd Guide	Rep1.BFP .RevGuide	Rep1.BFP2GFP .TopDonor.Recoded.5P.endPhos	Rep1.BFP2GFP.BotDonor.Recoded.5P.endPhos	NLS-nCas9-linker-SplintR-bpNLS	Yes

Example 2. Editing to convert BFP to GFP by Replacer 2

[00268] An experiment can be performed similar to Example 1 but adjusted to fit a Replacer 2 format. The ligases used here are T4 ligase, hLIG1(233-919), and hLIG1(119-919). The Replacer 2 gRNA contains a spacer, scaffold, and DBS. The gRNAs are delivered individually (1-sided Replacer 2) or in pairs (2-sided Replacer 2), and the gRNAs contain 2'-O-methyl 3'-phosphorothioate nucleotides at the first three and last three positions. The DNA donors include a FBS and a guide binding site (GBS) that can hybridize to the DBS. Some DNA donors contain SpPAM mutations and some DNA donors have phosphorothioate bonds at the first three and last three positions. Some DNA donors are recoded. The DNA donors are phosphorylated on the 5' end. The DNA donors are delivered as pairs in the Replacer 2 format. Some of the gRNAs and donor DNAs are annealed prior to LNP formulation. **Table 13** details this experiment. Sequences corresponding to the names in the table may be found herein.

Table 13

Con ditio n	Forwar d Guide	Reverse Guide	Top Donor	Bottom Donor	Ligase	Anneal both sides?
1	Rep2.B FP.Fwd Guide		Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker- hLIG1(233- 919)-bpNLS	Yes
2		Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker- hLIG1(233- 919)-bpNLS	Yes
3	Rep2.B FP.Fwd Guide	Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker- hLIG1(233- 919)-bpNLS	Yes
4	Rep2.B FP.Fwd Guide	Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker- hLIG1(233- 919)-bpNLS	No
5	Rep2.B FP.Fwd Guide	Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .Recoded.5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.Recoded.5P	NLS-nCas9- linker- hLIG1(233- 919)-bpNLS	Yes
6	Rep2.B FP.Fwd Guide	Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.5P	Rep2.BFP2GFP.B otDonor.5P	NLS-nCas9- linker- hLIG1(233- 919)-bpNLS	Yes
7	Rep2.B FP.Fwd Guide	Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .Recoded.5P.endPh os	Rep2.BFP2GFP.B otDonor.SpPAMm ut.Recoded.5P.end Phos	NLS-nCas9- linker- hLIG1(233- 919)-bpNLS	Yes
8	Rep2.B FP.Fwd Guide		Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker-T4LIG- bpNLS	Yes
9		Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker-T4LIG- bpNLS	Yes
10	Rep2.B FP.Fwd Guide	Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .Recoded.5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.Recoded.5P	NLS-nCas9- linker-T4LIG- bpNLS	Yes
11	Rep2.B FP.Fwd Guide		Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker- hLIG1(119- 919)-bpNLS	Yes
12		Rep2.BF P.RevG uide	Rep2.BFP2GFP.To pDonor.SpPAMmut .5P	Rep2.BFP2GFP.B otDonor.SpPAMm ut.5P	NLS-nCas9- linker- hLIG1(119- 919)-bpNLS	Yes

Condition	Forward Guide	Reverse Guide	Top Donor	Bottom Donor	Ligase	Anneal both sides?
13	Rep2.BFP.Fwd Guide	Rep2.BFP.Rev Guide	Rep2.BFP2GFP.TopDonor.SpPAMmut.Recoded.5P	Rep2.BFP2GFP.BotDonor.SpPAMmut.Recoded.5P	NLS-nCas9-linker-hLIG1(119-919)-bpNLS	Yes

Example 3. Editing to insert mGL in front of CBX1 by Replacer 2

[00269] An editing experiment can be performed to insert monomeric Green Lantern (mGL) in the genome of HEK293T cells in front of the CBX1 gene such that a fusion protein is formed that exhibits green fluorescence. This fluorescence can be detected by flow cytometry as in Examples 1 and 2. The experiment is conducted in a similar way to Example 2 except that the sequences of the gRNAs and DNA donors are different and enable insertion of mGL into the genome rather than insertion of a sequence that changes blue fluorescent protein (BFP) to green fluorescent protein (GFP). The DNA donors in Example 3 are longer than in Example 2 and are synthesized by GenScript. The DNA donors are phosphorylated on the 5' end. **Table 14** details this experiment. Sequences corresponding to the names in the table may be found herein.

Table 14

Condition	Forward Guide	Reverse Guide	Top Donor	Bottom Donor	Ligase	Anneal both sides?
1	Rep2.CBX1.FwdGuide		Rep2.mGL-CBX1.TopDonor.SpPAMmut.5P	Rep2.mGL-CBX1.BotDonor.SpPAMmut.5P	NLS-nCas9-linker-hLIG1(233-919)-bpNLS	Yes
2		Rep2.CBX1.RevGuide	Rep2.mGL-CBX1.TopDonor.SpPAMmut.5P	Rep2.mGL-CBX1.BotDonor.SpPAMmut.5P	NLS-nCas9-linker-hLIG1(233-919)-bpNLS	Yes
3	Rep2.CBX1.FwdGuide	Rep2.CBX1.RevGuide	Rep2.mGL-CBX1.TopDonor.SpPAMmut.5P	Rep2.mGL-CBX1.BotDonor.SpPAMmut.5P	NLS-nCas9-linker-hLIG1(233-919)-bpNLS	Yes
4	Rep2.CBX1.FwdGuide		Rep2.mGL-CBX1.TopDonor.SpPAMmut.5P	Rep2.mGL-CBX1.BotDonor.SpPAMmut.5P	NLS-nCas9-linker-T4LIG-bpNLS	Yes

Condition	Forward Guide	Reverse Guide	Top Donor	Bottom Donor	Ligase	Anneal both sides?
5		Rep2.CBX1 .RevGuide	Rep2.mGL- CBX1.TopDonor .SpPAMmut.5P	Rep2.mGL- CBX1.BotDonor.S pPAMmut.5P	NLS- nCas9- linker- T4LIG- bpNLS	Yes
6	Rep2.CBX1. FwdGuide	Rep2.CBX1 .RevGuide	Rep2.mGL- CBX1.TopDonor .SpPAMmut.5P	Rep2.mGL- CBX1.BotDonor.S pPAMmut.5P	NLS- nCas9- linker- T4LIG- bpNLS	Yes
7	Rep2.CBX1. FwdGuide		Rep2.mGL- CBX1.TopDonor .SpPAMmut.5P	Rep2.mGL- CBX1.BotDonor.S pPAMmut.5P	NLS- nCas9- linker- hLIG1(1 19-919)- bpNLS	Yes
8		Rep2.CBX1 .RevGuide	Rep2.mGL- CBX1.TopDonor .SpPAMmut.5P	Rep2.mGL- CBX1.BotDonor.S pPAMmut.5P	NLS- nCas9- linker- hLIG1(1 19-919)- bpNLS	Yes
9	Rep2.CBX1. FwdGuide	Rep2.CBX1 .RevGuide	Rep2.mGL- CBX1.TopDonor .SpPAMmut.5P	Rep2.mGL- CBX1.BotDonor.S pPAMmut.5P	NLS- nCas9- linker- hLIG1(1 19-919)- bpNLS	Yes

Example 4. Treatment of a genetic disease in a patient

[00270] A human patient with sickle cell disease comes to a physician for treatment. The patient is identified as having a hemoglobin gene mutation. Hematopoietic stem and progenitor cells are collected from the patient's peripheral blood. The cells are edited by contacting the cells' genomes with a nCas9-DNA ligase fusion protein, a gRNA, and a donor DNA that includes a corrected hemoglobin gene. The gRNA recruits the fusion protein to the gene mutation, and the nCas9 nicks the patient's DNA on one side flanking the mutation. The gRNA binds to a genomic flap generated by the nick, and to the donor DNA, and forms an RNA splint for the ligase to ligate the genomic flap to the donor DNA. Another fusion protein nicks the opposite strand of the mutated hemoglobin gene using a second gRNA on the other side of the mutation, and ligates the other side of the donor DNA. The mutated DNA is thus replaced with the donor DNA, and the cell with the donor DNA is transfused back into the patient, thus treating the genetic disease in the patient.

Example 5. Enhancing a crop

[00271] In a soybean plant, a germ cell is microinjected with an expression vector encoding an nCas9-DNA ligase fusion protein, and with a gRNA and donor DNA encoding an herbicide resistance gene. gRNA recruits the fusion protein to a suitable spot within the soybean genome which doesn't already include a gene. The nCas9 nicks the soybean's DNA on one side flanking the spot. The gRNA also recruits the donor DNA to bind to a genomic flap created by the nick, and the ligase seals the nick using the donor DNA itself as a splint. Another fusion protein nicks the opposite strand of the soybean's DNA on the other side flanking the spot, and ligates the other side of the donor DNA, thus integrating the herbicide resistance gene into the germ cell. The germ cell eventually produces a seed, and the seeds are harvested to grow herbicide resistant soybeans.

Example 6. *In Vitro* 1-Sided Replacer 2 using T4 Ligase

[00272] To demonstrate the usefulness of the components and methods described herein for editing nucleic acids, *in vitro* experiments were performed. The experiments in this example specifically assessed the feasibility of 1-sided Replacer 2. The experiments used a 100bp, 5'-Cy5-labeled double-stranded DNA (dsDNA) substrate (IDT) that corresponded to the blue fluorescent protein (BFP) target region (see examples 1 and 2), with the site of nicking located in the middle at base pair 50. 5'-phosphorylated dsDNA donors (IDT) containing a variable GBS, 13nt flap binding site (FBS), and a protospacer adjacent motif (PAM) mutation were used in conjunction with gRNAs (Agilent) containing the corresponding variable DBS. 5'-Cy5-labeled dsDNA substrate and 5'-phosphorylated dsDNA donor were separately annealed using complementary oligonucleotides by heating to 95C for 5min followed by slowly cooling to room temperature.

[00273] *In vitro* 1-sided Replacer 2 reactions were performed by first incubating gRNA (30nM final) and dsDNA donor (30nM final) with recombinant *S. pyogenes* nicking Cas9 (Cas9n; IDT; 30nM final) for 10min at room temperature, followed by the addition of T4 ligase (NEB; 200U final), ATP (1mM final), and 5'-Cy5-labeled dsDNA substrate (3nM Final). Reactions were carried out in the presence of NEB Buffer 3.1 (1x final) at 37C for 1hr (final volume of 10ul). Reactions were terminated by the addition of 0.5% SDS and 100ug/ml Proteinase K, and incubated at 37C for 30min. Reaction products were then combined with 2x formamide gel loading buffer (90% formamide; 10% glycerol; 0.01% bromophenol blue), denatured at 95 °C for 10min, and separated by denaturing urea PAGE gel (15% TBE-urea, 55 °C, 200 V). DNA products were visualized by Cy5 fluorescence signal using a LI-COR Odyssey CLx imager.

[00274] In addition to the intact 100bp 5'-Cy5-labeled dsDNA substrate, a nicked 5'-Cy5-labeled dsDNA substrate and a final ligation product were included as size controls. The nicked 5'-Cy5-labeled dsDNA control was annealed using two 50mers corresponding to the top strand oligo of the 100bp 5'-Cy5-labeled dsDNA substrate (a 5'-Cy5-labeled 50mer and a 5'-phosphorylated 50mer) and its complementary 100mer bottom strand oligo. The final ligation product control was annealed and ligated using the 5'-Cy5-labeled 50mer and the bottom 100mer from the nicked control along with the 150nt top strand donor oligo.

[00275] **Fig. 8A** illustrates an exemplary nicking and ligation pattern of an integrating nucleic acid. **Fig. 8B** illustrates an exemplary nucleic acid gel showing pattern associated with *In Vitro* 1-Sided Replacer 2 using 30nt GBS/DBS and Thermostable T4 Ligase. Using a 30nt GBS/DBS combination, a donor containing a PAM mutation, and a thermostable T4 ligase (Hi-T4, NEB), we were able to produce a final Replacer product (Lane 3) corresponding to the size of our control product (Lane 1). Replacer products were not detected in the absence of nicking Cas9 (Cas9n) (Lane 2), or in the absence of the bottom donor which serves as the splint (Lanes 4 & 5). **Fig. 8C** illustrates an exemplary nucleic acid gel showing pattern associated with *in vitro* 1-Sided Replacer 2 using Variable Length GBS/DBS Combinations and T4 Ligase. Using regular T4 ligase (NEB), we were to produce a final Replacer product corresponding to the size of the control when using multiple GBS/DBS combinations, including No GBS/DBS, 20nt GBS/DBS, and 30nt GBS/DBS. Additionally, in this experiment, recoded dsDNA donors containing PAM mutation were more efficient at producing final Replacer products compared to PAM mutant dsDNA donors that were not recoded. The results indicate that a DNA ligase may be used with an RNA-guided endonuclease to edit a target nucleic acid.

Example 7. Use of 1-sided Replacer 2 with nicking Cas9 and multiple DNA ligases in various coupling architectures in mammalian cells

[00276] Components used to edit a blue fluorescent protein (BFP) gene stably integrated into HEK293T cells were co-delivered by lipofectamine 2000 transfection. The components included a chemically synthesized guide RNA (SEQ ID NO: 166, mG*mC*mU*GAAGCACUGCACGCCAUGUUUUAGAGCUAGAAAUAGCAAGUUAAA AUAAGGCUAGUCCGUUAUCGACUUGAAAAAGUCGGACCGAGUCGGUCCAGCUGC GGUAUUGUGGmC*mG*mU) with 2'-O-methyl and phosphorothioate chemical modifications on the 5' and 3' ends, an integrating nucleic acid with a 5' phosphate end modification (SEQ ID NO: 167, /5Phos/cgtaTgtcagggtgtcacGAGgg), a splinting nucleic acid with locked nucleic acid and phosphorothioate modifications (SEQ ID NO: 169, +c*c*+CT+CG+TG+AC+CA+CC+CT+GA+CA+TA+CGGCGTGCAgtgcttACGCCA+CA+A

T+AC+CG+CA+G*C*+T), and either a single mRNA encoding nicking Cas9 fused to a ligase, or a pair of mRNAs encoding nicking Cas9 and a ligase.

[00277] The integrating nucleic acid and splinting nucleic acid were synthesized by Integrated DNA Technologies (IDT). All mRNAs corresponding to Cas9n (H840A) and all ligases are generated via in vitro transcription (IVT) reactions using the HiScribe T7 High Yield RNA Synthesis Kit (NEB). Coding sequences are cloned into an IVT vector that contains a single copy of the 5'UTR and two copies of the 3'UTR from the human beta globin gene, in addition to a 152nt polyA tail. Plasmid DNA containing coding sequences are linearized using an XbaI restriction site located immediately downstream of the polyA tail. Linearized plasmids are then purified via phenol:chloroform extraction followed by ethanol precipitation. mRNAs are produced via IVT reactions that contain N1-Methylpseudouridine-5'-Triphosphate (TriLink BioTech) in place of Uridine-Triphosphate, and capped co-transcriptionally with CleanCap Reagent AG (3' OMe) (TriLink BioTech). IVT reactions are incubated at 37°C for 2 hours, followed by DNase I digestion of the template DNA. Finally, mRNA products are purified using LiCl precipitation, quantified (Qubit Fluorometric Quantification; ThermoFisher), and checked for integrity by denaturing gel electrophoresis. "Ligase in trans" refers to Cas9 H840A nickase combined with T4 ligase fused to leucine zipper on its C terminus (T4-LZ, SEQ ID NO: 145). "LZ; C terminal Ligase" refers to Cas9 H840A nickase fused to a leucine zipper on its C terminus (nCas9-LZ, SEQ ID NO: 133) combined with a ligase fused to a leucine zipper on its N terminus for T4 (LZ-T4, SEQ ID NO: 142), SplintR (LZ-SplintR, SEQ ID NO: 141), or hLIG4(1-620) (LZ-hLIG4(1-620), SEQ ID NO: 146). "LZ; N terminal Ligase" refers to Cas9 H840A nickase fused to a leucine zipper on its N terminus (LZ-nCas9, SEQ ID NO: 147) combined with a ligase fused to a leucine zipper on its C terminus for T4 (T4-LZ, SEQ ID NO: 145), SplintR (SplintR-LZ, SEQ ID NO: 148), or hLIG4(1-620) (hLIG4(1-620)-LZ, SEQ ID NO: 149). "Fusion; C terminal Ligase" refers to Cas9 H840A nickase fused to a ligase with the ligase on the C terminus for T4 (nCas9-T4, SEQ ID NO: 131), SplintR (nCas9-SplintR, SEQ ID NO: 129), or hLIG4(1-620) (nCas9-hLIG4(1-620) SEQ ID NO: 150). "Fusion; N terminal Ligase" refers to Cas9 H840A nickase fused to a ligase with the ligase on the N terminus for T4 (T4-nCas9, SEQ ID NO: 151), SplintR (SplintR-nCas9, SEQ ID NO: 152), or hLIG4(1-620) (hLIG4(1-620)-nCas9, (SEQ ID NO: 153). The gRNA contained a spacer, scaffold, and donor binding site. The splinting integrating nucleic acid contained a guide binding site and a flap binding site. The ligating integrating nucleic acid and splinting nucleic acid were partially complementary.

[00278] The integrating nucleic acid and splinting nucleic acid were hybridized using an annealing reaction, then mixed with the guide RNA and mRNA and formulated with

lipofectamine 2000 in OptiMEM prior to delivery to the adherent HEK293 cells in 96-well plates. After 24-48 hours, the cells were detached with 0.05% Trypsin-EDTA and run through a flow cytometer to measure the percentage of cells expressing green fluorescent protein (GFP), indicating gene editing from BFP to GFP (**Fig. 9**). Gene editing was observed with T4, SplintR, and hLIG4(1-620) ligases when fused to nCas9, interacting with nCas9 through leucine zippers, or delivered in trans with no leucine zipper interaction.

[00279] The results here demonstrate the usefulness of using a DNA ligase with an RNA-guided endonuclease to edit a target nucleic acid in a cell. The experiments in this example specifically demonstrated the feasibility of including 1-sided Replacer 2 components to edit a target nucleic acid in a mammalian cell. This example shows the effectiveness of including a DNA ligase coupled through a heterodimerization domain (here, leucine zippers) to an RNA guided endonuclease (e.g. a nicking Cas9) in nucleic acid editing such as gene editing. This also shows nucleic acid editing is possible in mammalian cells with a DNA ligase fused to an RNA guided endonuclease (e.g. T4 ligase fused to Cas9 H840A nickase), and that nucleic acid editing can be achieved by delivering the DNA ligase and RNA guided endonuclease as separate non-coupled components.

Example 8. Use of 1-Sided Replacer 2 with nicking Cas9 and T4 DNA Ligase to make a variety of edits at multiple genomic targets

Components used to edit genomic targets in HEK293T cells were co-delivered by lipofectamine 2000 transfection. The components included a chemically synthesized guide with 2'-O-methyl and phosphorothioate chemical modifications on the 5' and 3' ends, an integrating nucleic acid with a 5' phosphate end modification, a splinting nucleic acid with locked nucleic acid and phosphorothioate modifications, an mRNA encoding nicking Cas9 (LZ-nCas9, SEQ ID NO: 147), and an mRNA encoding a ligase (T4-LZ, SEQ ID NO: 145). Target-specific guides, splinting and integrating nucleic acids are listed in Table 15. The integrating nucleic acid and splinting nucleic acid were synthesized by Integrated DNA Technologies (IDT) and both mRNAs were generated via in vitro transcription reactions using the methods described in Example 7. The gRNA contained a spacer, scaffold, and donor binding site. The splinting integrating nucleic acid contained a guide binding site and a flap binding site. The ligating integrating nucleic acid and splinting nucleic acid were partially complementary. The integrating nucleic acid and splinting nucleic acid were hybridized using an annealing reaction, then mixed with the guide RNA and mRNA and formulated with lipofectamine 2000 in OptiMEM prior to delivery to the adherent HEK293 cells in 96-well plates. After 24-48 hours, genomic DNA was extracted from the cells using QuickExtract and genomic targets were amplified using Q5 DNA Polymerase. The PCR program ran at 98C for 30 seconds, then 35

cycles of 98C for 5 seconds, 67C for 20 seconds, and 72C for 20 seconds, then finally 72C for 2 minutes. PCR primers are listed in **Table 15**. PCR products were cleaned up with ExoCIP treatment and submitted for next generation sequencing (NGS) by Azenta using their Amplicon-EZ service. Sequencing reads were merged and aligned to the amplicon of interest, and the percentage total reads that matched the intended edit was calculated (**Fig. 10**). This example shows the effectiveness of gene editing with 1-sided Replacer 2 in mammalian cells at a variety of genomic targets. The types of edits here include making a single point mutation (HEK3 F +5 G to T), a pair of point mutations (VEGFA R +5 G to T and +2 A to T, VEGFA F +5 G to T and +2 G to C, and AAVS1 R +5 G to T), or a trinucleotide insertion (HEK3 F CAC insertion and AAVS1 R CAC insertion) using 1-sided Replacer 2.

Table 15

Condition	Guide	Splint	Donor	PCR Primers
VEGFA R +5 G to T and +2 A to T	SEQ ID NO: 170 mC*mA*mC*CCCGG CUCUGGCUAAAGGU UUUAGAGCUAGAA AUAGCAAGUAAAA AUAAGGCUAGUCCG UUAUCGACUUGAA AAAGUCGGACCGAG UCGGUCCAGCUGCG GUAUUGUGGmC*mG *mU	SEQ ID NO: 174 +C*C*+TT +TC+CA+ AA+GC+C C+AT+TC +CA+TC+ ATtagccag agccggAC GCCA+C A+AT+AC +CG+CA+ G*C*+T	SEQ ID NO: 180 /5Phos/ATGA TGGAATGG GCTTTGGAA AGG	SEQ ID NO: 186 ACACTCTTTCCCTA CACGACGCTCTTCC GATCTTGCCGCTCA CTTTGATGTCT; SEQ ID NO: 187 GACTGGAGTTCAG ACGTGTGCTCTTCC GATCTGGGGAGAG GGACACACAGA
VEGFA F +5 G to T and +2 G to C	SEQ ID NO: 171 mG*mA*mU*GUCUG CAGGCCAGAUGAGU UUUAGAGCUAGAA AUAGCAAGUAAAA AUAAGGCUAGUCCG UUAUCGACUUGAA AAAGUCGGACCGAG UCGGUCCGUGGUUC CGGGCUGCAmU*mG *mA	SEQ ID NO: 175 +A*C*+A A+TG+TG +CC+AT+ CT+GG+A G+CA+CT +GAtctggc ctgcagaTC ATGC+A G+CC+CG +GA+AC+ C*A*+C	SEQ ID NO: 181 /5Phos/TCAG TGCTCCAGA TGGCACATT GT	SEQ ID NO: 223 ACACTCTTTCCCTA CACGACGCTCTTCC GATCTTGCCGCTCA CTTTGATGTCT; SEQ ID NO: 224 GACTGGAGTTCAG ACGTGTGCTCTTCC GATCTGGGGAGAG GGACACACAGA
HEK3 F CAC insertion	SEQ ID NO: 172 mG*mG*mC*ccagacug agcagugaGUUUUAG AGCUAGAAAUAAGC AAGUAAAAUAAG GCUAGUCCGUUAUC	SEQ ID NO: 176 +G*C*+T T+CC+TT +TC+CT+ CT+GC+C	SEQ ID NO: 182 /5Phos/gtgTG ATGGCAGA GGAAAGGA AGC	SEQ ID NO: 188 ACACTCTTTCCCTA CACGACGCTCTTCC GATCTccctggcctgggtc aatcc; SEQ ID NO: 189 GACTGGAGTTCAG

Condition	Guide	Splint	Donor	PCR Primers
	GACUUGAAAAAGU CGGACCGAGUCGGU CCGUGGUUCCGGGC UGCAmU*mG*mA	A+TC+Ac +accgtgctc agtctgTCA TGC+AG+ CC+CG+G A+AC+C* A*+C		ACGTGTGCTCTTCC GATCTctgaaggccaggt ccctc
HEK3 F +5 G to T	SEQ ID NO: 221 mG*mG*mC*ccagacug agcacgugaGUUUUAG AGCUAGAAAUAGC AAGUUAUUUAAG GCUAGUCCGUUAUC GACUUGAAAAAGU CGGACCGAGUCGGU CCGUGGUUCCGGGC UGCAmU*mG*mA	SEQ ID NO: 177 +A*G*+G G+CT+TC +CT+TT+ CC+TC+T G+CA+AT +CAcgtgct cagtctgTC ATGC+A G+CC+CG +GA+AC+ C*A*+C	SEQ ID NO: 183 /5Phos/TGAT TGCAGAGG AAAGGAAG CCCT	SEQ ID NO: 225 ACACTCTTTCCCTA CACGACGCTCTTCC GATCTccctggcctgggtc aatcc; SEQ ID NO: 226 GACTGGAGTTCAG ACGTGTGCTCTTCC GATCTctgaaggccaggt ccctc
AAVS1 R CAC insertion	SEQ ID NO: 173 mG*mC*mG*acuccugg aaguggccaGUUUUAG AGCUAGAAAUAGC AAGUUAUUUAAG GCUAGUCCGUUAUC GACUUGAAAAAGU CGGACCGAGUCGGU CCAGCUGCGGUAAU GUGGmC*mG*mU	SEQ ID NO: 178 +A*T*+T A+GC+AG +AA+GT+ GG+CC+C T+TG+Gc +accacttcc aggACGC CA+CA+A T+AC+CG +CA+G*C *+T	SEQ ID NO: 184 /5Phos/gtgCC AAGGGCCA CTTCTGCTA AT	SEQ ID NO: 190 ACACTCTTTCCCTA CACGACGCTCTTCC GATCTCGCCGGGA ACTGCCGCTGGC; SEQ ID NO: 191 GACTGGAGTTCAG ACGTGTGCTCTTCC GATCTGAGGAGGC CCTCATCTGGCG
AAVS1 R +5 G to T	SEQ ID NO: 222 mG*mC*mG*acuccugg aaguggccaGUUUUAG AGCUAGAAAUAGC AAGUUAUUUAAG GCUAGUCCGUUAUC GACUUGAAAAAGU CGGACCGAGUCGGU CCAGCUGCGGUAAU GUGGmC*mG*mU	SEQ ID NO: 179 +T*C*+C A+TT+AG +CA+GA+ AG+TG+G C+CA+TT +GGccactt ccaggACG CCA+CA+ AT+AC+C G+CA+G* C*+T	SEQ ID NO: 185 /5Phos/CCAA TGGCCACTT CTGCTAATG GA	SEQ ID NO: 227 ACACTCTTTCCCTA CACGACGCTCTTCC GATCTCGCCGGGA ACTGCCGCTGGC; SEQ ID NO: 228 GACTGGAGTTCAG ACGTGTGCTCTTCC GATCTGAGGAGGC CCTCATCTGGCG

Example 9. Use of 2-Sided Replacer 2 with nicking Cas9 and T4 DNA Ligase to make deletions and sequence replacements at multiple genomic targets

[00280] Components used to edit genomic targets in HEK293T cells were co-delivered by lipofectamine 2000 transfection. The components included two chemically synthesized guides with 2'-O-methyl and phosphorothioate chemical modifications on the 5' and 3' ends, two integrating nucleic acids with a 5' phosphate end modification, two splinting nucleic acids with locked nucleic acid and phosphorothioate modifications, an mRNA encoding nicking Cas9 (LZ-nCas9, SEQ ID NO: 147), and an mRNA encoding a ligase (T4-LZ, SEQ ID NO: 145). For both "VEGFA replacement of 175nt with attB" and "VEGFA 175nt deletion", the two guide RNAs used were VEGFA_R (SEQ ID NO: 170) and VEGFA_F (SEQ ID NO: 171). For both "AAVS1 replacement of 117nt with attB" and "AAVS1 117nt deletion", the two guide RNAs used were AAVS1_R (SEQ ID NO: 173) and AAVS1_F (SEQ ID NO: 192), mG*mC*mU*ggccccccaccgccccGUUUUAGAGCUAGAAAUAGCAAGUUAUUAAUAAGG CUAGUCCGUUAUCGACUUGAAAAAGUCGGACCGAGUCGGUCCGUGGUUCCGGGC UGCAmU*mG*mA). For "VEGFA replacement of 175nt with attB", the splinting nucleic acids used were SEQ ID NO: 193

(+g*g*+ag+ac+cg+cc+gt+cg+tc+ga+ca+ag+cctctggcctgcagaTCATGC+AG+CC+CG+GA+AC +C*A*+C) and SEQ ID NO: 194

(+g*g*+cg+gt+ct+cc+gt+cg+tc+ag+ga+tc+attagccagagccggACGCCA+CA+AT+AC+CG+CA +G*C*+T), and the integrating nucleic acids used were SEQ ID NO: 195

(/5Phos/ggcttgctgacgacggcggtctcc) and SEQ ID NO: 196 (/5Phos/atgacctgacgacggagaccgcc).

For "VEGFA 175nt deletion", the splinting nucleic acids used were SEQ ID NO: 197

(+C*C*+GT+CT+GC+AC+AC+CC+CG+GC+TC+TG+GC+TAAtctggcctgcagaTCATGC+AG+ CC+CG+GA+AC+C*A*+C) and SEQ ID NO: 198

(+G*C*+TC+AC+TT+TG+AT+GT+CT+GC+AG+GC+CA+GAtagccagagccggACGCCA+CA +AT+AC+CG+CA+G*C*+T), and the integrating nucleic acids used were SEQ ID NO: 199

(/5Phos/TAGCCAGAGCCGGGGTGTGCAGACGG) and SEQ ID NO: 200

(/5Phos/TCTGGCCTGCAGACATCAAAGTGAGC). For "AAVS1 replacement of 117nt with attB", the splinting nucleic acids used were SEQ ID NO: 201

(+g*g*+ag+ac+cg+cc+gt+cg+tc+ga+ca+ag+ccggcggtgggTCATGC+AG+CC+CG+GA+AC+C *A*+C) and SEQ ID NO: 202

(+g*g*+cg+gt+ct+cc+gt+cg+tc+ag+ga+tc+atccacttcaggACGCCA+CA+AT+AC+CG+CA+G *C*+T), and the integrating nucleic acids used were SEQ ID NO: 195 and SEQ ID NO: 196.

For "AAVS1 117nt deletion", the splinting nucleic acids used were SEQ ID NO: 203

(+C*G*+GG+GC+AC+AG+CG+AC+TC+CT+GG+AA+GT+GGggcggtgggTCATGC+AG+C

C+CG+GA+AC+C*A*+C) and SEQ ID NO: 204 (+G*G*+AA+CT+GC+CG+CT+GG+CC+CC+AC+CG+CCcacttccaggACGCCA+CA+AT+AC+CG+CA+G*C*+T), and the integrating nucleic acids used were SEQ ID NO: 205 (/5Phos/CCACTTCCAGGAGTCGCTGTGCCCCG) and SEQ ID NO: 206 (/5Phos/GGCGGTGGGGGGCCAGCGGCAGTTCC). The integrating nucleic acid and splinting nucleic acid were synthesized by Integrated DNA Technologies (IDT) and both mRNAs were generated via in vitro transcription reactions using the methods described in Example 7. The gRNA contained a spacer, scaffold, and donor binding site. The splinting integrating nucleic acid contained a guide binding site and a flap binding site. There were two pairs of ligating integrating nucleic acid and splinting nucleic acid, and each pair was partially complementary to each other. The integrating nucleic acid and splinting nucleic acid were hybridized using an annealing reaction, then mixed with the guide RNA and mRNA and formulated with lipofectamine 2000 in OptiMEM prior to delivery to the adherent HEK293 cells in 96-well plates. After 24-48 hours, genomic DNA was extracted from the cells using QuickExtract and genomic targets were amplified using Q5 DNA Polymerase. The PCR program ran at 98C for 30 seconds, then 35 cycles of 98C for 5 seconds, 67C for 20 seconds, and 72C for 20 seconds, then finally 72C for 2 minutes. PCR primers used for both “VEGFA replacement of 175nt with attB” and “VEGFA 175nt deletion” are SEQ ID NO: 186 and SEQ ID NO: 187. PCR primers used for both “AAVS1 replacement of 117nt with attB” and “AAVS1 117nt deletion” are SEQ ID NO: 190 and SEQ ID NO: 191. PCR products were cleaned up with ExoCIP treatment and submitted for next generation sequencing (NGS). Sequencing reads were merged and aligned to the amplicon of interest, and the percentage total reads that matched the intended edit was calculated (**Fig. 11**). This example shows that when Replacer 2 is delivered as 2 full sets of guide RNA, splint, and donor, it can delete an entire region of DNA between the nicking sites on each guide RNA, and optionally replace that region of DNA with a new DNA sequence. Since Replacer is making two separate flaps that can hybridize to each other here, this gene editing mechanism would not rely on the MMR pathway. After an attB sequence is inserted into a targeted site in the genome by Replacer, an entire synthetic gene could be inserted at that attB site if it is delivered with a Bxb1 integrase. Thus, the attB sequence replacement described here could be used for targeted insertion of large 1kb+ DNA fragments into the genome without double strand break or mismatch repair mediated gene editing.

Example 10. Use of 1-Sided Replacer 2 with nicking Cas9 and T4 DNA Ligase to integrate methylated DNA into a genomic target

[00281] Components used to edit genomic targets in HEK293T cells were co-delivered by lipofectamine 2000 transfection. The components included a chemically synthesized guide with 2'-O-methyl and phosphorothioate chemical modifications on the 5' and 3' ends (SEQ ID NO: 166), an integrating nucleic acid, a splinting nucleic acid, an mRNA encoding nicking Cas9 (LZ-nCas9, SEQ ID NO: 147), and an mRNA encoding a ligase (T4-LZ, SEQ ID NO: 145). Conditions with the "non-methylated donor" used an integrating nucleic acid with a 5' phosphate end modification (SEQ ID NO: 207, /5Phos/CGTATGTCAGGGTGGTCACG). Conditions with the "donor with all cytosines methylated" used an integrating nucleic acid with a 5' phosphate end modification and methylated cytosines (SEQ ID NO: 207, /5Phos//5Me-dC/gtaTgt/iMe-dC/agggtggt/iMe-dC/a/iMe-dC/G). Conditions under "Splint is LNA" used a splinting nucleic acid with locked nucleic acid and phosphorothioate modifications (SEQ ID NO: 208, +C*g*+tg+ac+ca+cc+ct+ga+cA+TA+CGGCGTGCAGtgcttACGCCA+CA+AT+AC+CG+CA+G*C*+T). Conditions under "Splint is OMe" used a splinting nucleic acid with locked nucleic acid, 2'-O-methyl, and phosphorothioate modifications (SEQ ID NO: 209, mC*g*mUgmacmcamccmctmgamcAmUAmCGGCGTGCAGtgcttACGCCA+CA+AT+AC+CG+CA+G*C*+T). The integrating nucleic acid and splinting nucleic acid were synthesized by Integrated DNA Technologies (IDT) and both mRNAs were generated via in vitro transcription reactions using the methods described in Example 7. The gRNA contained a spacer, scaffold, and donor binding site. The splinting integrating nucleic acid contained a guide binding site and a flap binding site. The ligating integrating nucleic acid and splinting nucleic acid were partially complementary. The integrating nucleic acid and splinting nucleic acid were hybridized using an annealing reaction, then mixed with the guide RNA and mRNA and formulated with lipofectamine 2000 in OptiMEM prior to delivery to the adherent HEK293 cells in 96-well plates. After 24-48 hours, the cells were detached with 0.05% Trypsin-EDTA and run through a flow cytometer to measure the percentage of cells expressing green fluorescent protein (GFP), indicating gene editing from BFP to GFP (**Fig. 12**). This example shows that methylated DNA can be used in the integrating nucleic acid and does not negatively impact editing efficiency under ideal conditions, when the splint has LNA bases. When the splint has OMe bases instead of LNAs and thus lower affinity to the donor, methylated DNA in the donor boosts efficiency, showing that DNA methylation can improve the system by stabilizing the nucleic acid components. A methylated donor could also be used to specifically introduce DNA methylation into the genome at functional epigenetic sites such as promoters to regulate gene expression. A

follow-up experiment could be conducted by performing bisulfite sequencing on the genomic region that Replacer is introducing methylated DNA into to confirm that epigenetic editing has occurred. If Replacer successfully introduces DNA methylation into this genomic region and it is believed that the region's methylation state controls gene expression, quantitative PCR could be conducted to confirm that a gene of interest has reduced mRNA expression after editing.

[00282] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

[00283] While the foregoing disclosure has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the disclosure. For example, all the techniques and apparatus described above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually and separately indicated to be incorporated by reference for all purposes.

WHAT IS CLAIMED IS:

1. A system, comprising:
an endonuclease coupled to a DNA ligase; and
an integrating nucleic acid configured to be ligated by the DNA ligase to a strand break generated by the endonuclease in a target nucleic acid.
2. The system of claim 1, wherein the coupling is covalent.
3. The system of claim 2, comprising a fusion protein comprising the endonuclease and the DNA ligase.
4. The system of claim 3, wherein the endonuclease is amino (N)-terminal relative to the DNA ligase within the fusion protein.
5. The system of claim 3, wherein the endonuclease is carboxy (C)-terminal relative to the DNA ligase within the fusion protein.
6. The system of claim 3, wherein the fusion protein comprises a linker comprising 1-100 amino acids connecting the endonuclease and the DNA ligase.
7. The system of claim 1, wherein the coupling is non-covalent.
8. The system of claim 7, wherein the endonuclease coupled to the DNA ligase comprises a first polypeptide comprising at least part of the endonuclease, and a second polypeptide comprising at least part of the DNA ligase, wherein the first and second polypeptides are non-covalently coupled.
9. The system of claim 8, wherein the first polypeptide comprises a first heterodimerization domain that binds a second heterodimerization domain, and wherein the second polypeptide comprises the second heterodimerization domain.
10. The system of claim 9, wherein the heterodimer domains comprise a leucine zipper, PDZ domain, streptavidin, streptavidin binding protein, foldon domain, hydrophobic moiety, or a functional binding fragment thereof.
11. The system of claim 8, wherein the first polypeptide comprises a first intein that binds a second intein, and wherein the second polypeptide comprises the second intein.
12. The system of claim 1, wherein the ligase comprises a hairpin binding motif, and wherein the endonuclease and the DNA ligase are coupled with a nucleic acid comprising a scaffold that binds to the endonuclease and a hairpin that binds to the hairpin binding motif.
13. The system of claim 12, wherein the hairpin binding motif comprises an MS2 coat protein (MCP) peptide, and wherein the hairpin comprises an MS2 hairpin.
14. The system of claim 1, wherein the endonuclease and the DNA ligase are coupled with a heterobifunctional molecule comprising an endonuclease binding domain and a DNA ligase binding domain.

15. The system of claim 14, wherein the heterobifunctional molecule comprises a small molecule.
16. A system comprising a cell containing an RNA-guided endonuclease and a DNA ligase, both of which are heterologous to the cell, and an integrating nucleic acid configured to be ligated by the DNA ligase to a strand break generated by the endonuclease in a target nucleic acid.
17. A system comprising a cell containing a heterologous RNA-guided endonuclease, an endogenous DNA ligase, and an integrating nucleic acid configured to be ligated by the DNA ligase to a strand break generated by the endonuclease in a target nucleic acid.
18. The system of any one of claims 1-17, wherein the endonuclease comprises an RNA-guided endonuclease.
19. The system of claim 18, wherein the endonuclease comprises a class II CRISPR/Cas endonuclease.
20. The system of claim 19, wherein the endonuclease comprises a Cas9 endonuclease.
21. The system of claim 18, wherein the endonuclease comprises a nickase.
22. The system of claim 21, wherein the endonuclease comprises a Cas9 nickase.
23. The system of any one of claims 1-17, wherein the DNA ligase ligates DNA strands base paired to a DNA splint.
24. The system of any one of claims 1-17, wherein the DNA ligase ligates DNA strands base paired to an RNA splint.
25. The system of any one of claims 1-17, wherein the endonuclease or the DNA ligase comprises a nuclear localization signal, chromatin modifying domain, cell penetrating peptide, or tag polypeptide.
26. The system of claim 18, further comprising a guide RNA.
27. One or more nucleic acids encoding the endonuclease or DNA ligase of any one of claims 1-16.
28. A cell comprising the system of any one of claims 1-15.
29. A system of nucleic acids comprising:
 - a. a guide nucleic acid comprising:
 - i. a spacer complementary to a region of a genomic locus of a genomic strand,
 - ii. a scaffold for complexing with an endonuclease,
 - iii. an optional donor binding site that is at least partially complementary to an integrating nucleic acid, and

- iv. a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus; and
 - b. an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by the endonuclease.
- 30. A system of nucleic acids comprising:
 - a. a guide nucleic acid comprising:
 - i. a spacer complementary to a region of a genomic locus of a genomic strand,
 - ii. a scaffold for complexing with an endonuclease, and
 - iii. an optional donor binding site that is at least partially complementary to a splinting nucleic acid;
 - b. an integrating nucleic acid comprising a 5' end to be ligated to a 3' terminus of the genomic strand generated by the endonuclease; and
 - c. a splinting nucleic acid comprising a flap binding site that is at least partially identical or complementary to a genomic flap at or adjacent to the genomic locus, and comprising an optional guide binding site that is at least partially complementary to a guide nucleic acid.
- 31. The system of claim 29 or 30, wherein the genomic strand is in a cell.
- 32. The system of claim 29 or 30, wherein the splinting nucleic acid further comprises a donor binding site that is at least partially identical or complementary to a portion of the integrating nucleic acid.
- 33. The system of any one of claims 29 or 30, wherein the guide nucleic acid comprises a sequence of linking nucleic acids between the scaffold and the donor binding site.
- 34. The system of any one of claims 29 or 30, wherein the guide nucleic acid, the integrating nucleic acid, or the splinting nucleic acid comprises a modified internucleoside linkage.
- 35. The system of claim 34, wherein the modified internucleoside linkage comprises a phosphorothioate linkage.
- 36. The system of claim 29 or 30, wherein the guide nucleic acid, the integrating nucleic acid, or the splinting nucleic acid comprises a modified nucleoside.
- 37. The system of claim 36, wherein the modified nucleoside comprises a locked nucleic acid (LNA), a 2'fluoro, a 2' O-alkyl, a methylated cytosine, an inverted thymidine, or a combination thereof.
- 38. The system of claim 29 or 30, wherein the endonuclease comprises an RNA-guided endonuclease.

- 39. The system of claim 38, wherein the endonuclease comprises a class II CRISPR/Cas endonuclease.
- 40. The system of claim 39, wherein the endonuclease comprises a Cas9 endonuclease.
- 41. The system of claim 38, wherein the endonuclease comprises a nickase.
- 42. The system of claim 41, wherein the endonuclease comprises a Cas9 nickase.