MODIFIED CRISPR-CAS EFFECTOR POLYPEPTIDES AND METHODS OF USE THEREOF

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/619,705 filed on January 10, 2024, and U.S. Provisional Patent Application No. 63/727,069 filed on December 2, 2024, each of which is incorporated by reference in its entirety.

REFERENCE TO A SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on January 8, 2025, is named 60555-702_601_SL. xml and is 132,814,069 bytes in size.

BACKGROUND

[0003] Precise delivery of molecular and cellular therapeutic payloads (e.g., gene-manipulating nucleic acid payloads) unlocks a wide range of therapeutic applications. The growing understanding of payload delivery promotes the application of various molecular and cellular therapeutics to selectively function on hitherto "undruggable" cells, proteins, transcripts, and genes, thus potentially broadening therapeutic targets. For example, several molecular and cellular therapeutics have been approved for clinical use such as the antisense oligonucleotide (ASO) mipomersen for familial hypercholesterolemia, and the splice-switching oligonucleotide (SSO) eteplirsen for Duchenne muscular dystrophy. Molecular and cellular therapeutics also include siRNA delivery, with a recent shift toward mRNA delivery. The shift to mRNA payloads opens a myriad of therapeutic applications. These range from expression of antigens of choice as prophylactic viral and bacterial vaccines as well as therapeutic cancer vaccine applications, supplementation of missing proteins for enzyme replacement therapies, and expression of gene editing machinery, such as CRISPR, to gene edit aberrant natively expressed proteins.

[0004] In the world of delivering molecular and cellular therapeutic payloads (e.g., RNA-based therapeutics) into cells, most, if not all, roads can eventually lead to the endosomal escape abyss. Delivery of molecular and cellular therapeutics often involves internalization via endocytosis followed by trafficking through various membrane-bound vesicular compartments. For example, endocytosed molecular and cellular therapeutics can be transferred to early endosomes, which can mature into late endosomes and eventually into lysosomes. The development of delivery modalities, such as LNPs, has addressed some of the challenges related to inefficient delivery. However, despite

recent developments in the field, a commonly overlooked aspect regards the very limited release of the payloads into the cytoplasm. For efficient delivery, in some situations, the payloads would need to be released into the cytosol before the maturation of late endosomes to lysosomes where the majority of the foreign materials are either degraded enzymatically or recycled outside of the target cells, with only a very limited amount of payload released into the cytoplasm. The release of the payload prior to lysosomal maturation can be a crucial stage for efficient delivery and is known as endosomal escape. This process can be inefficient and is considered a bottleneck. For example, large, hydrophilic, negatively-charged properties often prevent molecular and cellular therapeutic payloads (e.g., nucleic acid payloads) from passively diffusing across the lipid bilayers. Thus, escaping from the endosome and releasing payloads into the cytoplasm in a non-toxic manner is a critical technical problem, resulting in inefficient access of therapeutic payloads to the sites of action in the nucleus or cytosol of cells. This problem has meant that large payload doses are often administered to attain therapeutic effects, risking drug-related toxicities. One approach to overcoming this challenge has been the use of complex delivery systems such as cationic lipid or polymer nanoparticles, however these approaches often suffer from toxicity and biodistribution problems associated with the delivery system itself. Therefore, there remains a need for the development of alternative strategies to enhance the delivery of molecular and cellular therapeutic payloads to their intracellular targets.

[0005] Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas systems comprise a CRISPR-associated (Cas) effector polypeptide and a guide nucleic acid. Such CRISPR-Cas systems can bind to and modify a targeted nucleic acid. The programmable nature of these CRISPR-Cas effector systems has facilitated their use as a versatile technology for use in, e.g., gene editing. Many therapeutic approaches involve delivery of a nucleic acid vector or messenger RNA encoding a CRISPR-Cas protein, rather than delivery of the CRISPR-Cas protein per se. Delivery of nucleic acid vectors encoding a CRISPR-Cas protein, however, has certain disadvantages, namely possible insertional mutagenesis and the possibility of off-target gene editing. There is a need in the art for modified CRISPR-Cas effector polypeptides that are readily delivered as the proteins per se.

SUMMARY

[0006] The compositions and methods described herein can overcome the drawbacks of the currently available approaches for payload delivery to sites of action in the nucleus and/or cytosol of cells. Provided herein, are compositions, methods, compositions, and kits useful for delivering therapeutic payloads to intracellular targets.

[0007] Provided herein, in some aspects, are compositions comprising a fusion polypeptide comprising: (a) an effector polypeptide; and (b) a targeting moiety that specifically binds to an

extracellular domain of a receptor expressed by a target cell, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody. In some embodiments, the fusion polypeptide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated. In some embodiments, the fusion polypeptide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

[0008] Provided herein, in some aspects, are compositions comprising a fusion polypeptide comprising: (a) an effector polypeptide; and (b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the fusion polypeptide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated. In some embodiments, the fusion polypeptide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

[0009] Provided herein, in some aspects, are compositions comprising a fusion polypeptide comprising: (a) an effector polypeptide; and (b) a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

[0010] Provided herein, in some aspects, are compositions comprising a fusion polypeptide comprising: (a) an effector polypeptide; and (b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 336-349**. In some embodiments, the fusion polypeptide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell. In some embodiments, the targeting moiety is not an antibody or an antigen binding domain of an antibody.

[0011] Provided herein, in some aspects, are compositions comprising a fusion polypeptide comprising: (a) an effector polypeptide; and (b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 152-335**.

[0012] Provided herein, in some aspects, are compositions comprising a protein complex, the protein complex comprising: (a) the fusion polypeptide of the composition of claim 7 or 8, and (b) an endosomal escape peptide, wherein the endosomal escape peptide is bound to the fusion polypeptide, and wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

[0013] In some embodiments, the fusion polypeptide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell. In some embodiments, the endosomal escape peptide is bound to a pH responsive polypeptide domain. In some embodiments, the targeting moiety binds to an extracellular domain of transferrin receptor protein 1 (TfR1), IGF1R, IGF2R, Dnmt3a, or Nephrin. In some embodiments, the targeting moiety comprises a polypeptide sequence with less than 150 amino acids. In some embodiments, the targeting moiety comprises a polypeptide sequence with less than 100 amino acids. In some embodiments, the targeting moiety comprises a polypeptide sequence with more than 50 amino acids. In some embodiments, the targeting moiety comprises a polypeptide sequence with more than 60 amino acids. In some embodiments, the targeting moiety comprises a polypeptide sequence with more than 70 amino acids. In some embodiments, the targeting moiety comprises a polypeptide sequence with 80 amino acids or more. In some embodiments, the targeting moiety comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the pH responsive polypeptide domain contains one or more amino acids that are protonated at a pH of 6.8 or less, and wherein the one or more amino acids that are protonated at a pH of 6.8 or less are not protonated at a pH of 7.4. In some embodiments, the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 6.6 or less, and wherein the one or more amino acids that are protonated at a pH of 6.6 or less are not protonated at a pH of 7.4. In some embodiments, the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 6 or less, and wherein the one or more amino acids that are protonated at a pH of 6 or less are not protonated at a pH of 7.4. In some embodiments, the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5.5 or less, and wherein the one or more amino acids that are protonated at a pH of 5.5 or less are not protonated at a pH of 7.4. In some embodiments, the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5 or less, and wherein the one or more amino acids that are protonated at a pH of 5 or less are not protonated at a pH of 7.4. In some embodiments, the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polypeptide is present within an endosome. In some

embodiments, the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polypeptide is present within a lysosome. In some embodiments, the pH responsive polypeptide domain comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 336-349**. In some embodiments, at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH

responsive polypeptide domain at a pH of 7.4. In some embodiments, the endosomal escape peptide is released from the protein complex when the protein complex is present within an endosome. In some embodiments, the endosomal escape peptide is released from the protein complex when the protein complex is present within a lysosome. In some embodiments, the endosomal escape peptide comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 152-335**. In some embodiments, the pH responsive polypeptide domain is capable of binding to two or more endosomal escape peptides at pH 7.4. In some embodiments, the protein complex comprises two or more copies of the endosomal escape peptide. In some embodiments, the fusion polypeptide further comprises a linker linking the effector polypeptide and one or more other peptide sequences of the fusion polypeptide. In some embodiments, the linker is a proteolytically cleavable linker. In some embodiments, the fusion polypeptide further comprises one or more nuclear localization sequences (NLSs). In some embodiments, the one or more NLSs comprise the amino acid sequence K(K/R)X(K/R), and where X is any amino acid. In some embodiments, the one or more NLSs comprise the amino acid sequence PKKKRKV. In some embodiments, the one or more NLSs are at the N-terminus of the effector polypeptide. In some embodiments, the effector polypeptide is a CRISPR-Cas effector polypeptide. In some embodiments, the CRISPR-Cas effector polypeptide is a Type II CRISPR-Cas effector polypeptide, a Type V CRISPR-Cas effector polypeptide, or a Type VI CRISPR-Cas effector polypeptide. In some embodiments, the CRISPR-Cas effector polypeptide is a SpyCas9 or a variant thereof. In some embodiments, the CRISPR-Cas effector polypeptide is GeoCas9 or a variant thereof. In some embodiments, the CRISPR-Cas effector polypeptide is catalytically active. In some embodiments, the CRISPR-Cas effector polypeptide exhibits reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide. In some embodiments, the CRISPR-Cas effector polypeptide is catalytically inactive. In some embodiments, the fusion polypeptide further comprises at least one additional heterologous polypeptide. In some embodiments, the at least one additional heterologous polypeptide is a deaminase, a base editor, a reverse transcriptase, a transcription modulator, or an epigenetic modulator.

[0014] Provided herein, in some aspects, are compositions comprising: (a) a fusion polypeptide of the compositions disclosed herein, or a nucleic acid comprising a nucleotide sequence encoding the fusion polypeptide and/or endosomal escape polypeptide; and (b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid. In some embodiments, the guide nucleic acid is a single-molecule guide nucleic acid. In some embodiments, the compositions further comprise a donor nucleic acid.

[0015] Provided herein, in some aspects, are nucleic acids comprising a nucleotide sequence encoding the fusion polypeptide of the compositions disclosed herein. In some embodiments, the nucleotide sequence is operably linked to a transcriptional control element, optionally wherein the transcriptional control element is a promoter.

[0016] Provided herein, in some aspects, are recombinant expression vectors comprising the nucleic acids disclosed herein.

[0017] Provided herein, in some aspects, are cells comprising the compositions disclosed herein, or the nucleic acids disclosed herein, or the recombinant expression vector disclosed herein. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is *in vitro*. In some embodiments, the cell is *in vivo*.

[0018] Provided herein, in some aspects, are methods of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) the compositions disclosed herein; and (b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

[0019] Provided herein, in some aspects, are methods of modifying a target nucleic acid in a eukaryotic cell, the method comprising introducing into the eukaryotic cell: (a) the compositions disclosed herein; and (b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid. In some embodiments, the methods further comprise introducing into the cell a donor nucleic acid. In some embodiments, the cell is *in vitro*. In some embodiments, the cell is *in vivo*. In some embodiments, the cell is a mammalian cell, an insect cell, an avian cell, a reptile cell, an amphibian cell, an arachnid cell, a protozoan cell, or a plant cell. In some embodiments, the target nucleic acid is selected from: double stranded DNA, single stranded DNA, RNA, genomic DNA, and extrachromosomal DNA. In some embodiments, the modifying comprises genome editing.

INCORPORATION BY REFERENCE

[0020] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by

reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

FIG. 1 shows exemplary configurations of the fusion polypeptides and fusion polynucleotides of the present disclosure.

DETAILED DESCRIPTION

[0022] Before the present invention is further described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0023] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0024] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0025] It may be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “an endosomal escape peptide” includes a plurality of such polypeptides and reference to “the CRISPR-Cas effector polypeptide” includes reference to one or more CRISPR-Cas effector polypeptides and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as “solely,” “only” and the like in connection with the recitation of claim elements, or use of a “negative” limitation.

[0026] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the invention are specifically embraced by the present invention and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present invention and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

[0027] The publications discussed herein do not constitute an admission of the art, but is used for helping with the understanding of the technology described.

COMPOSITIONS

[0028] The present disclosure provides compositions. Compositions provided herein can comprise a polymer (such as a polynucleotide, a polypeptide, etc.), a fusion polymer (such as a fusion polypeptide, a fusion polynucleotide, etc.), or any combination thereof. In some embodiments, the fusion polymer comprises a polymer and one or more heterologous polypeptides.

Polymer

[0029] Disclosed herein in some embodiments are compositions comprising polymers. A "polymer" as disclosed herein can refer to any class of natural or engineered substances composed of large molecules (e.g., macromolecules), which are multiples of monomers (e.g., simpler chemical units). In some embodiments, a polymer can comprise a polynucleotide, a polypeptide, or a combination thereof. In some embodiments, a polymer can be a natural polymer, an engineered polymer, or any combination thereof. In some embodiments, an engineered polymer can comprise an engineered polynucleotide, an engineered polypeptide, or a combination thereof. In some embodiments, an "engineered polymer", "engineered polynucleotide", or "engineered polypeptide" as referred to herein can comprise a composition that is not present in nature. In some embodiments, a composition that is not present in nature can comprise a composition that was produced artificially by humans. In some embodiments an engineered polymer can comprise a sequence that is not present in nature, a combination of components that are not present together in nature, a composition that has been altered from a composition that is otherwise present in nature, or a combination thereof.

Polypeptide

[0030] Disclosed herein in some embodiments are polymers comprising polypeptides. The terms "peptide," "polypeptide," and "protein" can be used interchangeably, and can refer to a polymeric

form of amino acids of any length, which can include genetically coded and non-genetically coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. The term includes fusion proteins, including, but not limited to, fusion proteins with a heterologous amino acid sequence. In some embodiments a protein or polypeptide can refer to a polymeric form of amino acids, and no limitation is placed on the maximum number of amino acids that can comprise a protein's or peptide's sequence. In some embodiments a polypeptide can comprise any peptide or protein comprising two or more amino acids joined to each other by peptide bonds. In some embodiments a polypeptide can comprise short chains, such as a peptide, an oligopeptide or an oligomers. In some embodiments a polypeptide can comprise longer chains, such as a protein, of which there are many types. In some embodiments a polypeptide can comprise a biologically active fragment, a substantially homologous polypeptide, an oligopeptide, a homodimer, a heterodimer, a variant of a polypeptide, a modified polypeptide, a derivative, an analog, or a fusion protein, among others. In some embodiments a polypeptide can comprise a natural peptide, a recombinant peptide, or a combination thereof. In some embodiments, a polypeptide can comprise an engineered polypeptide. An engineered polypeptide can refer to a synthetically constructed polypeptide. For example, a polypeptide can be synthesized by chemical synthesis, enzymatic synthesis, or any combination thereof. In some embodiments, the polypeptide is a synthetic polypeptide.

[0031] Polypeptides (e.g., proteins) can exist in different forms based on the number and arrangement of their subunits. These forms can be a single subunit (monomer), two subunits (dimer), or multiple subunits (multimeric proteins). In some cases, the polypeptide(s) disclosed herein can comprise one or more subunits (e.g., the polypeptide can be monomeric, dimeric, or multimeric). Incorporation of a modifications in the polypeptide can modulate the activity and/or functionality of the polypeptide. For example, incorporation of an amino acid at a particular position of the polypeptide can modulate the folding or the function of the polypeptide. In some cases, incorporation of the modifications can occur within one or more subunits of a multimeric polypeptide. In some cases, incorporation of the modifications at a particular position of the polypeptide can render the polypeptide functional, wherein the particular amino acid and modified amino acid have at least one difference in their chemical structures. Such difference can modulate the folding or the function of the polypeptide when the particular amino acid or the modified amino acid is incorporated at the particular position of the polypeptide. In other cases, incorporation of the modified amino acid at the particular position of the polypeptide can render the polypeptide non-functional. In some cases, incorporation of a particular amino acid in place of the modified amino acid at a particular position of the polypeptide can render the polypeptide non-functional, wherein the particular amino acid and

modified amino acid have at least one difference in their chemical structures. Similar considerations apply to any polypeptides described herein (for example, an engineered component). The identity of the modified amino acid have, the particular amino acid, and/or the particular position of the amino acid can depend on the identity of the polypeptide.

[0032] Polypeptides as described herein also include polypeptides having various amino acid additions, deletions, or substitutions relative to the native amino acid sequence of a polypeptide of the present disclosure. In some embodiments, polypeptides that are homologs of a polypeptide of the present disclosure contain non-conservative changes of certain amino acids relative to the native sequence of a polypeptide of the present disclosure. In some embodiments, polypeptides that are homologs of a polypeptide of the present disclosure contain conservative changes of certain amino acids relative to the native sequence of a polypeptide of the present disclosure, and thus may be referred to as conservatively modified variants. A conservatively modified variant may include individual substitutions, deletions or additions to a polypeptide sequence which result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well-known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure. The following eight groups contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, e.g., Creighton, Proteins (1984)). A modification of an amino acid to produce a chemically similar amino acid may be referred to as an analogous amino acid.

Amino acid modifications

[0033] In some embodiments, the polypeptides disclosed herein can comprise any type of amino acid molecule (e.g., standard amino acids, such as the 20 amino acids commonly found in proteins and used by the genetic code during protein synthesis, nonstandard amino acids, such as chemically modified standard amino acids, amino acids that occur in living organisms but are typically not found in proteins, etc.). In some embodiments, the polypeptide comprises one or more nonstandard amino acid, such as hydroxyproline, hydroxylysine, gamma-carboxyglutamate, desmosine, isodesmosine, citrulline, ornithine, beta-alanine, and selenocysteine. In some embodiments, the polypeptide comprises one or more amino acid modifications. Amino acid modifications can refer to post-translational modifications, such as modifications that can alter the structure and/or function of a polypeptide by changing the properties (e.g., chemical) of an amino acid. Post-translational

modifications can occur on the amino acid side chains of a polypeptide, at a C-terminus of the polypeptide, at an N-terminus of the polypeptide, or any combination thereof. Amino acid modifications can include, but are not limited to, acetylation, oxidation, hydroxylation, methylation, phosphorylation, amidation, carboxylation, and dehydroalanine. Amino acid modifications can impact a polypeptide's function, stability, and/or cellular localization.

[0034] In some cases, the polypeptides disclosed herein comprise an amino acid side chain modification. Amino acid side chain modifications can refer to phosphorylation, glycosylation, oxidative modification, proteolytic cleavage, alkene modifications, chemical modifiers, or any combination thereof. For example, a carbohydrate molecule can be incorporated into an amino acid side chain to improve stability and promote folding. In some cases, the polypeptides disclosed herein comprise a C-terminus modification. Modifications in the C-terminus of a polypeptide can include, but are not limited to, amidation (e.g., adding an amide group), prenylation (e.g., adding a lipid anchor), glycosylation (e.g., adding a sugar molecule), phosphorylation, methylation, and the addition of a GPI (glycosylphosphatidylinositol) anchor. For example, a GPI anchor can be incorporated into a C-terminus of a polypeptide to promote cellular localization of the polypeptide to specific cellular membranes. Modifications in the N-terminus of a polypeptide can include, but are not limited to, removal of the initiator methionine, addition of small chemical groups (e.g., acetylation, propionylation, methylation, myristoylation, palmitoylation, and/or ubiquitylation), and/or attachment of lipid anchors, such as myristoyl or palmitoyl groups. For example, an acetyl group can be added to an N-terminal amino group of a polypeptide to impact protein stability and/or protein-protein interactions.

Endosomal Escape Peptide

[0035] In some aspects, provided herein are polypeptides that can facilitate release of payloads (e.g., molecular and cellular therapeutics such as the fusion polypeptides disclosed herein or the effector polypeptides of the fusion polypeptides disclosed herein) into the cytoplasm. For example, endocytosed molecular and cellular therapeutics (such as payloads internalized into a cell via endocytosis) can be transferred to early endosomes, which can mature into late endosomes and eventually into lysosomes. For efficient delivery, the payloads must be released into the cytosol before the maturation of late endosomes to lysosomes where majority of the payload can be either degraded enzymatically or recycled outside of the target cells, with only a very limited amount of payload released into the cytoplasm (e.g., endosomal escape). In some embodiments, the polypeptide that can facilitate release of payloads into cytoplasm is an endosomal escape peptide. An endosomal escape peptide (EEP) can refer to a peptide that can disrupt the endosomal membrane. For example, the EEP can disrupt the endosomal membrane thus, promoting the escape of a payload (e.g., a fusion

polypeptide disclosed herein) from the endosome and releasing the payload into the cytoplasm, resulting in efficient access of therapeutic payloads to the sites of action in the nucleus or cytosol of cells. For example, an ASO delivered to a cell and internalized by endocytosis, can be released from the endosome by the EEP which disrupts the endosomal membrane. An EEP can be used to deliver molecules that are otherwise unable to pass through the cell membrane by themselves, such as proteins, and nucleic acids.

[0036] In some embodiments, the endosomal escape peptide comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 152-335**. In some embodiments, the endosomal escape peptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of **SEQ ID NOs: 152-335**. In some embodiments, the endosomal escape peptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 152-335**. In some embodiments, the endosomal escape peptide comprises an amino acid sequence to any one of **SEQ ID NOs: 152-335**. In some embodiments, the endosomal escape peptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 152-335**. In some embodiments, the endosomal escape peptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 152-335**. **TABLE 4** lists the sequences of **SEQ ID NOs: 152-335**.

Effector polypeptides

[0037] "Effector polypeptides" can refer to a polypeptide (e.g., zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), CRISPR/Cas, etc.) that directly interact with other molecules (e.g., RNA, DNA, polypeptides, etc.) to regulate gene expression (e.g., by inhibiting translation, cleaving DNA, causing degradation of a target mRNA, etc.), essentially acting as "effectors" to influence cellular processes. In some embodiments, the polypeptide can be

configured for polypeptide-mediated gene expression alteration. Gene expression alteration can include, but is not limited to, changing the level of activity of a gene, such as turning it up (e.g., increasing expression) or down (e.g., decreasing expression) through various mechanisms. Gene expression alterations can also refer to changes in gene regulation mechanisms, such as translation, transcription, epigenetics, post-transcription, post-translation, or any combination thereof. For example, in some cases, the polypeptide can be configured for polypeptide-mediated translation disruption, such as CRISPR-associated (Cas) proteins which can be used to activate or repress translation by binding to the 5' untranslated regions (UTRs) of messenger RNAs (mRNAs). Gene expression can be altered in a number of ways including, epigenetic alterations, DNA methylation, genome editing, site-directed mutagenesis, or any combination thereof. For example, in some cases, the polypeptide can be configured for polypeptide-mediated DNA cleavage, such as ZFNs which can lead to insertions, deletions, inversions, or gene disruption following DNA cleavage. In some embodiments, the effector polypeptide is an enzyme.

[0038] In some embodiments, the effector polypeptide comprises a nuclease. In some embodiments, the polypeptide comprises dTALEs (e.g., transcription activator-like effectors that are highly target specific and can distinguish 1–2 nucleotide differences in a target sequence), TALENs (e.g., a type of transcription activator-like effector (TALE) nuclease that can target any desired sequence within a genome, without restrictions on PAM sites), or ZFNs (e.g., a restriction enzyme that cuts DNA at a specific sequence in a genome). In some embodiments, the effector polypeptide is an endonuclease. In some embodiments, the effector polypeptide is a transposon-associated RNA-guided endonucleases. In some embodiments, the polypeptide comprises TnpB (e.g., a transposon-encoded protein that can be programmed to cleave DNA in the presence of a transposon-adjacent motif (TAM)), Fanzor (e.g., an RNA-guided DNA-cutting enzyme that can be reprogrammed to target specific sites in the genome), or Cas (e.g., CRISPR/Cas). In some embodiments, the polypeptide comprises a Cas enzyme.

[0039] Cas enzymes along with their associated Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) guide ribonucleic acids (RNAs) appear to be a pervasive (~45% of bacteria, ~84% of archaea) component of prokaryotic immune systems, serving to protect such microorganisms against non-self nucleic acids, such as infectious viruses and plasmids by CRISPR-RNA guided nucleic acid cleavage. While the deoxyribonucleic acid (DNA) elements encoding CRISPR RNA elements may be relatively conserved in structure and length, their CRISPR-associated (Cas) proteins are highly diverse, containing a wide variety of nucleic acid-interacting domains. While CRISPR DNA elements have been observed as early as 1987, the programmable endonuclease cleavage ability of CRISPR/Cas complexes has only been recognized relatively

recently, leading to the use of recombinant CRISPR/Cas systems in diverse DNA manipulation and gene editing applications. Owing to the utility of these enzymes, they are being repurposed for a wide variety of biotechnology, gene editing, and therapeutic applications. In some embodiments, the engineered polypeptide can comprise at least one CRISPR-Cas effector polypeptide. CRISPR-Cas effector polypeptides include, but are not limited to, Type I CRISPR-Cas effector polypeptides, Type II CRISPR-Cas effector polypeptides, Type III CRISPR Cas effector polypeptides, Type IV CRISPR-Cas effector polypeptides, Type V CRISPR Cas effector polypeptides, and Type VI CRISPR-Cas effector polypeptides. In some embodiments, the CRISPR-Cas effector polypeptide is a Type II CRISPR-Cas effector polypeptide, a Type V CRISPR-Cas effector polypeptide, or a Type VI CRISPR-Cas effector polypeptide.

[0040] In some cases, the CRISPR-Cas effector polypeptide is a type II CRISPR-Cas effector polypeptide. Type II CRISPR-Cas systems are considered the simplest in terms of components. In Type II CRISPR-Cas systems, the processing of the CRISPR array into mature crRNAs does not require the presence of a special endonuclease subunit, but rather a small trans-encoded crRNA (tracrRNA) with a region complementary to the array repeat sequence; the tracrRNA interacts with both its corresponding effector nuclease (e.g. Cas9) and the repeat sequence to form a precursor dsRNA structure, which is cleaved by endogenous RNAse III to generate a mature effector enzyme loaded with both tracrRNA and crRNA. Cas II nucleases are DNA nucleases. Type II effectors generally exhibit a structure comprising a RuvC-like endonuclease domain that adopts the RNase H fold with an unrelated HNH nuclease domain inserted within the folds of the RuvC-like nuclease domain. The RuvC-like domain is responsible for the cleavage of the target (e.g., crRNA complementary) DNA strand, while the HNH domain is responsible for cleavage of the displaced DNA strand. In some cases, the type II CRISPR-Cas effector polypeptide is a Cas9 polypeptide, e.g., *Staphylococcus aureus* Cas9, *Streptococcus pyogenes* Cas9 (SpCas9), etc. In some cases, the CRISPR-Cas effector polypeptide is a variant of a wild-type SpCas9 and comprises one or more of the following substitutions: A61R, L1111R, A1322R, D1135L, S1136W, G1218K, E1219Q, N1317R, R1333P, R1335A, and T1337R. In some cases, the CRISPR-Cas effector polypeptide is an SpG polypeptide or a SpRY polypeptide; see, e.g., Walton et al. (2020) *Science* 368:290, and WO 2019/051097. For example, a suitable CRISPR-Cas effector polypeptide is an SpCas9 polypeptide includes D1135V, R1135Q, and T1137R substitutions, relative to wild-type SpCas9. As another example, a suitable CRISPR-Cas effector polypeptide is an SpCas9 polypeptide includes D1135V, R1335Q, T1337R, and G1218R substitutions, relative to wild-type SpCas9. As another example, a suitable CRISPR-Cas effector polypeptide is an SpCas9 polypeptide includes D1135L, S1136W, G1218K, E1219Q, R1335A, and T1337R substitutions, relative to wild-type SpCas9. As another

example, a suitable CRISPR-Cas effector polypeptide is an SpCas9 polypeptide includes L1111R, A1322R, D1135L, S1136W, G1218K, E1219Q, R1335A, and T1337R substitutions, relative to wild-type SpCas9. As another example, a suitable CRISPR-Cas effector polypeptide is an SpCas9 polypeptide includes A61R, L1111R, A1322R, D1135L, S1136W, G1218K, E1219Q, N1317R, R1333P, R1335A, and T1337R substitutions, relative to wild-type SpCas9. In some cases, the CRISPR-Cas effector polypeptide is a SpyCas9 or a variant thereof.

[0041] In some cases, one or more surface Cys residues of a CRISPR-Cas effector polypeptide are substituted. As one example, in some cases, a Cys at position 80 of the SpCas9 amino acid sequence or a corresponding Cys in another CRISPR-Cas effector polypeptide is substituted with an amino acid other than a Cys. For example, in some cases, a Cys at position 80 of the SpCas9 amino acid sequence, or a corresponding Cys in another CRISPR-Cas effector polypeptide is substituted with a Ser. As another example, in some cases, a Cys at position 574 of the SpCas9 amino acid sequence, or a corresponding Cys in another CRISPR-Cas effector polypeptide is substituted with an amino acid other than a Cys.

[0042] In some cases, the CRISPR-Cas effector polypeptide is a type V CRISPR-Cas effector polypeptide, e.g., a Cas12a, a Cas12b, a Cas12c, a Cas12d, or a Cas12e polypeptide. In some cases, the CRISPR-Cas effector polypeptide is a type VI CRISPR-Cas effector polypeptide, e.g., a Cas13a polypeptide, a Cas13b polypeptide, a Cas13c polypeptide, or a Cas13d polypeptide. In some cases, the CRISPR-Cas effector polypeptide is a Cas14 polypeptide. In some cases, the CRISPR-Cas effector polypeptide is a Cas14a polypeptide, a Cas14b polypeptide, or a Cas14c polypeptide. In some cases, the CRISPR-Cas effector polypeptide is a Cas7-11 polypeptide; see, e.g., Özcan et al. (2021) *Nature* 597:720. In some cases, the CRISPR-Cas effector polypeptide is a CRISPRi polypeptide; see, e.g., Qi et al. (2013) *Cell* 152:1173; and Jensen et al. (2021) *Genome Research* doi:10.1101/gr.275607.121. In some cases, the CRISPR-Cas effector polypeptide is a CRISPRa polypeptide; see, e.g., Jensen et al. (2021) *Genome Research* doi:10.1101/gr.275607.121; and Breinig et al. (2019) *Nature Methods* 16:51. In some cases, the CRISPR-Cas effector polypeptide is a CRISPRoff polypeptide. See, e.g., Nuñez et al. (2021) *Cell* 184:2503. In some cases, the CRISPR-Cas effector polypeptide is a nickase. In some cases, the CRISPR-Cas effector polypeptide exhibits reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide.

[0043] As noted above, in some cases, a CRISPR-Cas effector polypeptide is a Type VI CRISPR-Cas effector polypeptide, such as a Cas13a polypeptide, a Cas13b polypeptide, a Cas13c polypeptide, a Cas13d polypeptide, a Cas13e polypeptide, a Cas13f polypeptide, a Cas13X polypeptide, or a Cas13Y polypeptide. As noted above, in some cases, a CRISPR-Cas effector polypeptide is a Type III CRISPR-Cas effector polypeptide. A suitable Type III CRISPR-Cas

effector polypeptide is a Cas7-11 polypeptide (see, e.g., Özcan et al. (2021) *Nature* 597:710). In some cases, a CRISPR-Cas effector polypeptide is an RNA-binding CRISPR-Cas effector polypeptide. For example, in some cases, the CRISPR-Cas effector polypeptide is an RFx Cas13d polypeptide. As another example, in some cases, the CRISPR-Cas effector polypeptide is a dRfxCas13d polypeptide. As another example, in some cases, the CRISPR-Cas effector polypeptide is a DjCas13d polypeptide. As another example, in some cases, the CRISPR-Cas effector polypeptide is a PspCas13b polypeptide, e.g., *Prevotella sp. P5-125* Cas13b. As another example, in some cases, the CRISPR-Cas effector polypeptide is a dPspCas13b polypeptide In some cases, a Cas13 polypeptide (e.g., a Cas13a polypeptide, a Cas13b polypeptide, a Cas13c polypeptide, a Cas13d polypeptide, a Cas13e polypeptide, a Cas13f polypeptide, a Cas13X polypeptide, or a Cas13Y polypeptide) comprises two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains, each comprising a HEPN motif, where each HEPN motif is RXXXXH, RXXXXXH, or RXXXXXXH, where X is any amino acid.

[0044] In some cases, the CRISPR-Cas effector polypeptide is a variant that exhibits reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide. For example, where the CRISPR-Cas effector polypeptide (e.g., a Cas13 polypeptide) comprises a HEPN domain, the CRISPR-Cas effector polypeptide can comprise a substitution of the Arg and/or the His in the HEPN motif, where such a mutation reduces the catalytic activity of the CRISPR-Cas effector polypeptide. In some cases, the RNA-binding CRISPR-Cas effector polypeptide binds, but does not cleave, a target RNA. In some cases, a CRISPR-Cas effector polypeptide is a variant that, when complexed with a guide nucleic acid, binds but does not cleave a target DNA.

[0045] As another example, in some cases, the CRISPR-Cas effector polypeptide is a Cas7-11 polypeptide (e.g., a DiCas7-11 polypeptide). See, e.g., Özcan et al. (2021) *Nature* 597:720. In some cases, the Cas7-11 polypeptide is a variant that exhibits reduced catalytic activity compared to a wild-type Cas7-11 effector polypeptide. For example, in some cases, the variant Cas7-11 polypeptide comprises a substitution of one or more of D177, D429, D654, D758, E959, and D998. For example, a variant Cas7-11 can have an Asp at position 177, an Ala at position 429, an Ala at position 654, an Asp at position 758, a Glu at position E959, and an Asp at position 998. In some cases, a variant Cas7-11 comprises a D429A substitution.

[0046] In some embodiments, as an alternative to SpyCas9, other Cas9 enzymes, like GeoCas9 that offer a number of features over SpyCas9 can be employed. Because of the lineage of GeoCas9 from thermophiles, there is no exposure to GeoCas9 in humans and therefore no preexisting antibodies exist in humans to GeoCas9. This expands the possible reach of an RNP therapeutic in terms of patient numbers. In addition, GeoCas9 RNPs have a higher melting temp than SpyCas9 RNPs (72 vs

40°C) leading to a more stable RNP complex that does not release guide RNA after complexation. This can be beneficial for circulation in blood where the guide RNA can disassemble and be degraded by nucleases. GeoCas9 RNPs are not only more stable in mouse and human plasma, but also at lower pH ranges and have an expanded window of endosomal escape, allowing GeoCas9 to escape from early and late endosomes. Unlike GeoCas9, SpyCas9 can have significant loss in activity after exposure to pH 6.0 and below.

[0047] In some embodiments, the CRISPR-Cas effector polypeptide is a GeoCas9 or variant thereof. In some embodiments, the CRISPR-Cas effector polypeptide comprises 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to GeoCas9. In some embodiments, the CRISPR-Cas effector polypeptide comprises at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to a GeoCas9. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to a GeoCas9. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to a GeoCas9. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to a GeoCas9**.** In some embodiments, the CRISPR-Cas effector polypeptide is a GeoCas9.

[0048] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 1-11**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of **SEQ ID NOs: 1-11**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at

most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 1-11**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to any one of **SEQ ID NOs: 1-11**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 1-11.** In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 1-11. TABLE 1** below lists the sequences of **SEQ ID NOs: 1-11**.

[0049] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 1**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 1**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 1**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 1**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 1**.

[0050] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 2**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 2**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 2**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 2**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 2**.

[0051] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 3**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 3**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 3**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 3**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 3**.

[0052] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 4**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 4**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 4**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 4**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 4**.

[0053] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 5**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 5**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 5**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 5**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16,

at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 5**.

[0054] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 6**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 6**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 6**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 6**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 6**.

[0055] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 7**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 7**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 7**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 7**. In some embodiments, the

CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 7**.

[0056] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 8**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 8**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 8**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 8**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 8**.

[0057] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 9**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 9**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 9**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at

least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 9**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 9**.

[0058] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 10**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 10**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 10**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 10**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 10**.

[0059] In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 11**. In some embodiments, the CRISPR-Cas effector polypeptide comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 11**. In some embodiments, the CRISPR-Cas effector polypeptide comprises 100% sequence identity to **SEQ ID NO: 11**. In some embodiments,

the CRISPR-Cas effector polypeptide comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to **SEQ ID NO: 11**. In some embodiments, the CRISPR-Cas effector polypeptide comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to **SEQ ID NO: 11**.

Table 1. Exemplary polypeptide amino acid sequences

SEQ ID NO.	Amino acid sequence	Organism
1	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKIL TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNAS LGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF ANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGI LQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFE KNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQIS EFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	*Streptococcus pyogenes* Cas9
2	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK RGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQK LSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEK YVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQ SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEEL RSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKK KPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIE NAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNL SLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFIL SPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLN NPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSK ISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVD	*Staphylococcus aureus* Cas9

	TRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNK GYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEI ETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKD DKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKL IMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAH LDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKEN YYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNND LLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNL YEVKSKKHPQIIKKG	
3	MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKK AKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFK SAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDN GIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSII YRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKT SEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGI NEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVT TMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLT DLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAK YLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLA QISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSED KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKE NKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHT KNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRY NSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYS KGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPA KEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANK FNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIG NDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEI AKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVF KDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVT GFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAA KGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGH GECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVN GNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNL VIKNEEYFEFVQNRNN	*Francisella tularensis* Cas12a
4	TQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELK PIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQAT YRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVT TTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKF KENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQT QIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIP LFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNE LNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAK EKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPT TLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLE MEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNG AILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDA AKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEP KKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPS SQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDF AKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRM AHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALL PNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYL KEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNRE KERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLN FGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQ LTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHES RKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFE KNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGI	*Acidaminococcus sp. BV3L6* Cas12a

	VFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSP VRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDL KLQNGISNQDWLAYIQELRN	
5	MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKG VKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEI NLRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTA FTGFFDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHE VQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKI KGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEVL EVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFG EWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQ EYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKND AVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYDILLK VDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATIL RYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLP KVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDS ISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVD KLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLS GGAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYKDK RFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNL LYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQ NWTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVK VEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKS MSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMY VPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFD WEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSL MLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADAN GAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH	Lachnospiraceae bacterium ND2006 (LbCas12a)
6	MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQEN LYRRSPNGDGEQECDKTAEECKAELLERLRARQVENGHRGPAGSDDEL LQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAG NKPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPL MRVYTDSEMSSVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWN QRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLE SKEQTAHYVTGRALRGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRN TRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLNHAKMF ATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKV ENGVAREVDDVTVPISMSEQLDNLLPRDPNEPIALYFRDYGAEQHFTGE FGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPY AAVFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSV DLGLRTSASISVFRVARKDELKPNSKGRVPFFFPIKGNDNLVAVHERSQL LKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRER SWAKLIEQPVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAV YESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDVVGGNSIEQIE YLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKK LADRIIMEALGYVYALDERGKGKWVAKYPPCQLILLEELSEYQFNNDR PPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDARTG APGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTG EGEIFVSPFSAEEGDFHQIHADLNAAQNLQQRLWSDFDISQIRLRCDWG EVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFA QEKLSEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMV NQRIEGYLVKQIRSRVPLQDSACENTGDI	AacCas12b
7	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPR RLARSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQL RVEALDRKLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIE ENRAILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIF SKQREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKR APKATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQAFQKNKIT YHDIRTLLHLPDDTYFKGIVYDRGESRKQNENIRFLELDAYHQIRKAVD KVYGKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPN	GeoCas9

	LANKVYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACER AGYTFTGPKKKQKTMLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSP VSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNPT GHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDS YTNKVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFSKKK RDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQ KVYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKV TAFYQRREQNKELAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNL GNYDDQKLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQT VVKTKLSEIKLDASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQE PLYKPKKNGEPGPVIRTVKIIDTKNQVIPLNDGKTVAYNSNIVRVDVFEK DGKYYCVPVYTMDIMKGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYP NDLIRIELPREKTVKTAAGEEINVKDVFVYYKTIDSANGGLELISHDHRF SLRGVGSRTLKRFEKYQVDVLGNIYKVRGEKRVGLASSAHSKPGKTIRP LQSTRD	
8	MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKE KIDNNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDD FLETEEVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQ ENEEEIEIDIRDEYTNKTLNDCSIILRIIENDELETKKSIYEIFKNINMSLYKI IEKIIENETEKVFENRYYEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEI LGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIADFVIKELEF WNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDK IVKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKKGNCDTEIFGIFK KHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKVRLKKM EKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDF SRLHAKEELDLELITFFASTNMELNKIFSRENINNDENIDFFGGDREKNY VLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISKE RDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDI KISEENNNDIKYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYV NKELYKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIENYYKNAQI SASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIK DINDNKTYERITVKTSDKTIVINDDFEYIISIFALLNSNAVINKIRNRFFAT SVWLNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFIQKMKEIEK DFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKKLEKIVIFDD ETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKSKILCR IIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLF LNIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLN DKLNGYSKEYKEKYIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYK KIRDLVEFNYLNKIESYLIDINWKLAIQMARFERDMHYIVNGLRELGIIK LSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKFEKICYGFGIDL SENSEINKPENESIRNYISHFYIVRNPFADYSIAEQIDRVSNLLSYSTRYNN STYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSVLELE SYNSDYIKNLIIELLTKIENTNDTL	Leptotrichia shahii (WP_018451595.1) Cas13a
9	MTTTMKISIEFLEPFRMTKWQESTRRNKNNKEFVRGQAFARWHRNKK DNTKGRPYITGTLLRSAVIRSAENLLTLSDGKISEKTCCPGKFDTEDKDR LLQLRQRSTLRWTDKNPCPDNAETYCPFCELLGRSGNDGKKAEKKDW RFRIHFGNLSLPGKPDFDGPKAIGSQRVLNRV**D**FKSGKAHDFFKAYEVD HTRFPRFEGEITIDNKVSAEARKLLCDSLKFTDRLCGALCVIRFDEYTPA ADSGKQTENVQAEPNANLAEKTAEQIISILDDNKKTEYTRLLADAIRSLR RSSKLVAGLPKDHDGKDDHYLWDIGKKKKDENSVTIRQILTTSADTKE LKNAGKWREFCEKLGEALYLKSKDMSGGLKITRRILGDAEFHGKPDRL EKSRSVSIGSVLKETVVCGELVAKTPFFFGAIDEDAKQT**D**LQVLLTPDN KYRLPRSAVRGILRRDLQTYFDSPCNAELGGRPCMCKTCRIMRGITVMD ARSEYNAPPEIRHRTRINPFTGTVAEGALFNMEVAPEGIVFPFQLRYRGS EDGLPDALKTVLKWWAEGQAFMSGAASTGKGRFRMENAKYETLDLS DENQRNDYLKNWGWRDEKGLEELKKRLNSGLPEPGNYRDPKWHEINV SIEMASPFINGDPIRAAVDKRGT**D**VVTFVKYKAEGEEAKPVCAYKAESF RGVIRSAVARIHMEDGVPLTELTHSDCECLLCQIFGSEYEAGKIRFEDLV FESDPEPVTFDHVAIDRFTGGAADKKKF**D**DSPLPGSPARPLMLKGSFWI RRDVLEDEEYCKALGKALADVNNGLYPLGGKSAIGYGQVKSLGIKGD	DiCas7-11

	DKRISRLMNPAFDETDVAVPEKPKTDAEVRIEAEKVYYPHYFVEPHKK VEREEKPCGHQKFHEGRLTGKIRCKLITKTPLIVPDTSNDDFFRPADKEA RKEKDEYHKSYAFFRLHKQIMIPGSELRGMVSSVYETVTNSCFRIFDET KRLSWRMDADHQNVLQDFLPGRVTADGKHIQKFSETARVPFYDKTQK HFDILDEQEIAGEKPVRMWVKRFIKRLSLVDPAKHPQKKQDNKWKRRK EGIATFIEQKNGSYYFNVVTNNGCTSFHLWHKPDNFDQEKLEGIQNGEK LDCWVRDSRYQKAFQEIPENDPDGWECKEGYLHVVGPSKVEFSDKKG DVINNFQGTLPSVPNDWKTIRTNDFKNRKRKNEPVFCCEDDKGNYYTM AKYCETFFFDLKENEEYEIPEKARIKYKELLRVYNNNPQAVPESVFQSR VARENVEKLKSGDLVYFKHNEKYVEDIVPVRISRTVDDRMIGKRMSAD LRPCHGDWVEDGDLSALNAYPEKRLLLRHPKGLCPACRLFGTGSYKGR VRFGFASLENDPEWLIPGKNPGDPFHGGPVMLSLLERPRPTWSIPGSDN KFKVPGRKFYVHHHAWKTIKDGNHPTTGKAIEQSPNNRTVEALAGGNS FSFEIAFENLKEWELGLLIHSLQLEKGLAHKLGMAKSMGFGSVEIDVES VRLRKDWKQWRNGNSEIPNWLGKGFAKLKEWFRDELDFIENLKKLLW FPEGDQAPRVCYPMLRKKDDPNGNSGYEELKDGEFKKEDRQKKLTTP WTPWA	
10	IEKKKSFAKGMGVKSTLVSGSKVYMTTFAEGSDARLEKIVEGDSIRSVN EGEAFSAEMADKNAGYKIGNAKFSHPKGYAVVANNPLYTGPVQQDML GLKETLEKRYFGESADGNDNICIQVIHNILDIEKILAEYITNAAYAVNNIS GLDKDIIGFGKFSTVYTYDEFKDPEHHRAAFNNNDKLINAIKAQYDEFD NFLDNPRLGYFGQAFFSKEGRNYIINYGNECYDILALLSGLRHWVVHNN EEESRISRTWLYNLDKNLDNEYISTLNYLYDRITNELTNSFSKNSAANVN YIAETLGINPAEFAEQYFRFSIMKEQKNLGFNITKLREVMLDRKDMSEIR KNHKVFDSIRTKVYTMMDFVIYRYYIEEDAKVAAANKSLPDNEKSLSE KDIFVINLRGSFNDDQKDALYYDEANRIWRKLENIMHNIKEFRGNKTRE YKKKDAPRLPRILPAGRDVSAFSKLMYALTMFLDGKEINDLLTTLINKF DNIQSFLKVMPLIGVNAKFVEEYAFFKDSAKIADELRLIKSFARMGEPIA DARRAMYIDAIRILGTNLSYDELKALADTFSLDENGNKLKKGKHGMRN FIINNVISNKRFHYLIRYGDPAHLHEIAKNEAVVKFVLGRIADIQKKQGQ NGKNQIDRYYETCIGKDKGKSVSEKVDALTKIITGMNYDQFDKKRSVIE DTGRENAEREKFKKIISLYLTVIYHILKNIVNINARYVIGFHCVERDAQLY KEKGYDINLKKLEEKGFSSVTKLCAGIDETAPDKRKDVEKEMAERAKE SIDSLESANPKLYANYIKYSDEKKAEEFTRQINREKAKTALNAYLRNTK WNVIIREDLLRIDNKTCTLFRNKAVHLEVARYVHAYINDIAEVNSYFQL YHYIMQRIIMNERYEKSSGKVSEYFDAVNDEKKYNDRLLKLLCVPFGY CIPRFKNLSIEALFDRNEAAKFDKEKKKVSGNS	RfxCas13d
11	NQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQNENNENLWFHPV MSHLYNAKNGYDKQPEKTMFIIERLQSYFPFLKIMAENQREYSNGKYK QNRVEVNSNDIFEVLKRAFGVLKMYRDLTNHYKTYEEKLNDGCEFLTS TEQPLSGMINNYYTVALRNMNERYGYKTEDLAFIQDKRFKFVKDAYG KKKSQVNTGFFLSLQDYNGDTQKKLHLSGVGIALLICLFLDKQYINIFLS RLPIFSSYNAQSEERRIIIRSFGINSIKLPKDRIHSEKSNKSVAMDMLNEVK RCPDELFTTLSAEKQSRFRIISDDHNEVLMKRSSDRFVPLLLQYIDYGKL FDHIRFHVNMGKLRYLLKADKTCIDGQTRVRVIEQPLNGFGRLEEAET MRKQENGTFGNSGIRIRDFENMKRDDANPANYPYIVDTYTHYILENNK VEMFINDKEDSAPLLPVIEDDRYVVKTIPSCRMSTLEIPAMAFHMFLFGS KKTEKLIVDVHNRYKRLFQAMQKEEVTAENIASFGIAESDLPQKILDLIS GNAHGKDVDAFIRLTVDDMLTDTERRIKRFKDDRKSIRSADNKMGKRG FKQISTGKLADFLAKDIVLFQPSVNDGENKITGLNYRIMQSAIAVYDSGD DYEAKQQFKLMFEKARLIGKGTTEPHPFLYKVFARSIPANAVEFYERYL IERKFYLTGLSNEIKKGNRVDVPFIRRDQNKWKTPAMKTLGRIYSEDLP VELPRQMFDNEIKSHLKSLPQMEGIDFNNANVTYLIAEYMKRVLDDDF QTFYQWNRNYRYMDMLKGEYDRKGSLQHCFTSVEEREGLWKERASR TERYRKQASNKIRSNRQMRNASSEEIETILDKRLSNSRNEYQKSEKVIRR YRVQDALLFLLAKKTLTELADFDGERFKLKEIMPDAEKGILSEIMPMSF TFEKGGKKYTITSEGMKLKNYGDFFVLASDKRIGNLLELVGSDIVSKEDI MEEFNKYDQCRPEISSIVFNLEKWAFDTYPELSARVDREEKVDFKSILKI LLNNKNINKEQSDILRKIRNAFDHNNYPDKGVVEIKALPEIAMSIKKAFG EYAIMK	PspCas13b

[0060] In some embodiments, the CRISPR-Cas effector polypeptide is catalytically inactive. In some embodiments, the CRISPR-Cas effector polypeptide is catalytically active. In some embodiments, the CRISPR-Cas effector polypeptide exhibits reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide. In some embodiments, the CRISPR-Cas effector polypeptide exhibits at least 5%, at least 10%, 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 more reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide. In some embodiments, the CRISPR-Cas effector polypeptide exhibits at most at most 5%, at most 10%, at most 15%, 20%, at most 25%, at most 30%, at most 35%, at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, or at most 95% reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide.

Engineered guide ribonucleic acid

[0061] The present disclosure provides a composition comprising a fusion polypeptide of the present disclosure. In some cases, a composition of the present disclosure comprises: i) a polypeptide of the present disclosure (e.g., a fusion polypeptide); and ii) a guide nucleic acid. In some cases, a composition of the present disclosure comprises: i) a polypeptide of the present disclosure (e.g., a fusion polypeptide); ii) a guide nucleic acid (e.g., an engineered guide ribonucleic acid structure); and iii) a donor nucleic acid.

[0062] In one aspect, the present disclosure provides a composition comprising a CRISPR-Cas effector polypeptide and an engineered guide ribonucleic acid structure. The engineered guide ribonucleic acid structure may comprise a guide ribonucleic acid sequence. The guide ribonucleic acid sequence may be configured to hybridize to a target deoxyribonucleic acid sequence. The engineered guide ribonucleic acid structure may comprise a tracr ribonucleic acid sequence. The tracr ribonucleic acid sequence may be configured to bind to the CRISPR-Cas effector polypeptide. The engineered guide ribonucleic acid structure may lack a tracr ribonucleic acid sequence. For example, the engineered guide ribonucleic acid structure may function as a single molecule capable of binding to a target DNA without the need for a separate tracrRNA component, such as a single-guide RNA structure characteristic of a type V CRISPR system (e.g., Cas12a (Cpf1) which utilizes a single guide RNA structure that lacks a tracrRNA component). In some embodiments, the composition comprises an engineered guide ribonucleic acid structure or a nucleic acid encoding the engineered guide ribonucleic acid structure. In some embodiments, the engineered guide ribonucleic acid structure is configured to form a complex with the CRISPR-Cas effector polypeptide. In some embodiments, the engineered guide ribonucleic acid structure comprises a guide ribonucleic acid

sequence configured to hybridize to a target nucleic acid sequence; and a tracr ribonucleic acid sequence. In some embodiments, the engineered guide ribonucleic acid structure comprises at least two ribonucleic acid polynucleotides; or a single ribonucleic acid polynucleotide comprising the guide ribonucleic acid sequence and the tracr ribonucleic acid sequence. In some embodiments, the tracr ribonucleic acid sequence is configured to bind to the CRISPR-Cas effector polypeptide. In some embodiments, the guide ribonucleic acid sequence is complementary to a eukaryotic, a fungal, a plant, a mammalian, or a human genomic sequence. In some embodiments, the engineered guide ribonucleic acid structure comprises a ribonucleic acid sequence comprising a stem and a loop, wherein said stem comprises at least 12 pairs of ribonucleotides. In some embodiments, the engineered guide ribonucleic acid structure further comprises a second stem and a second loop, wherein the second stem comprises at least 5 pairs of ribonucleotides. In some embodiments, the engineered guide ribonucleic acid structure further comprises a ribonucleic acid structure comprising at least two hairpins.

[0063] Guide nucleic acids (e.g., an engineered guide ribonucleic acid structure) that form a complex with a CRISPR-Cas effector polypeptide are well known in the art. A guide RNA (e.g., an engineered guide ribonucleic acid structure) can be said to include two segments, a first segment (referred to herein as a "targeting segment"); and a second segment (referred to herein as a "protein-binding segment"). By "segment" it is meant a segment/section/region of a molecule, e.g., a contiguous stretch of nucleotides in a nucleic acid molecule. A segment can also mean a region/section of a complex such that a segment may comprise regions of more than one molecule. The "targeting segment" is also referred to herein as a "variable region" of a guide RNA. The "protein-binding segment" is also referred to herein as a "constant region" of a guide RNA.

[0064] In many instances, the targeting segment and the protein-binding segment are heterologous to one another. In some cases, the targeting segment comprises a nucleotide sequence that is complementary to a nucleotide sequence in a eukaryotic target nucleic acid.

[0065] The first segment (targeting segment) of a guide RNA includes a nucleotide sequence (a guide sequence) that is complementary to (and therefore hybridizes with) a specific sequence (a target site) within a target nucleic acid (e.g., a target ssRNA, a target ssDNA, the complementary strand of a double stranded target DNA, etc.). The protein-binding segment (or "protein-binding sequence") interacts with (binds to) a CRISPR/Cas effector polypeptide. The protein-binding segment of a guide RNA includes two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (dsRNA duplex). Site-specific binding and/or cleavage of a target nucleic acid (e.g., genomic DNA) can occur at locations (e.g., target sequence of

a target locus) determined by base-pairing complementarity between the guide RNA (the guide sequence of the guide RNA) and the target nucleic acid.

[0066] A guide RNA and a CRISPR/Cas effector polypeptide form a complex (e.g., bind via non-covalent interactions). The guide RNA provides target specificity to the complex by including a targeting segment, which includes a guide sequence (a nucleotide sequence that is complementary to a sequence of a target nucleic acid). The CRISPR/Cas effector polypeptide of the complex provides the site-specific activity (e.g., cleavage activity or an activity provided by the CRISPR/Cas effector polypeptide when the CRISPR/Cas effector polypeptide is a CRISPR/Cas effector polypeptide fusion polypeptide, i.e., has a fusion partner). In other words, the CRISPR/Cas effector polypeptide is guided to a target nucleic acid sequence (e.g. a target sequence in a chromosomal nucleic acid, e.g., a chromosome; a target sequence in an extrachromosomal nucleic acid, e.g. an episomal nucleic acid, a minicircle, an ssRNA, an ssDNA, etc.; a target sequence in a mitochondrial nucleic acid; a target sequence in a chloroplast nucleic acid; a target sequence in a plasmid; a target sequence in a viral nucleic acid; etc.) by virtue of its association with the guide RNA.

[0067] The "guide sequence" also referred to as the "targeting sequence" or "spacer sequence" of a guide RNA can be modified so that the guide RNA can target a CRISPR/Cas effector polypeptide to any desired sequence of any desired target nucleic acid, with the exception that the protospacer adjacent motif (PAM) sequence can be taken into account. Thus, for example, a guide RNA can have a targeting segment with a sequence (a guide sequence) that has complementarity with (e.g., can hybridize to) a sequence in a nucleic acid in a eukaryotic cell, e.g., a viral nucleic acid, a eukaryotic nucleic acid (e.g., a eukaryotic chromosome, chromosomal sequence, a eukaryotic RNA, etc.), and the like.

[0068] In some cases, a guide RNA includes two separate nucleic acid molecules and is referred to herein as a "dual guide RNA", a "double-molecule guide RNA", or a "two-molecule guide RNA" a "dual guide RNA", or a "dgRNA." In some cases, a guide RNA is a single-molecule RNA and is referred to as a "single guide RNA", or simply "sgRNA."

[0069] The targeting segment can have a length of 7 or more nucleotides (nt) (e.g., 8 or more, 9 or more, 10 or more, 12 or more, 15 or more, 20 or more, 25 or more, 30 or more, or 40 or more nucleotides). In some cases, the targeting segment can have a length of from 7 to 100 nucleotides (nt) (e.g., from 7 to 80 nt, from 7 to 60 nt, from 7 to 40 nt, from 7 to 30 nt, from 7 to 25 nt, from 7 to 22 nt, from 7 to 20 nt, from 7 to 18 nt, from 8 to 80 nt, from 8 to 60 nt, from 8 to 40 nt, from 8 to 30 nt, from 8 to 25 nt, from 8 to 22 nt, from 8 to 20 nt, from 8 to 18 nt, from 10 to 100 nt, from 10 to 80 nt, from 10 to 60 nt, from 10 to 40 nt, from 10 to 30 nt, from 10 to 25 nt, from 10 to 22 nt, from 10 to 20 nt, from 10 to 18 nt, from 12 to 100 nt, from 12 to 80 nt, from 12 to 60 nt, from 12 to 40 nt, from 12

to 30 nt, from 12 to 25 nt, from 12 to 22 nt, from 12 to 20 nt, from 12 to 18 nt, from 14 to 100 nt, from 14 to 80 nt, from 14 to 60 nt, from 14 to 40 nt, from 14 to 30 nt, from 14 to 25 nt, from 14 to 22 nt, from 14 to 20 nt, from 14 to 18 nt, from 16 to 100 nt, from 16 to 80 nt, from 16 to 60 nt, from 16 to 40 nt, from 16 to 30 nt, from 16 to 25 nt, from 16 to 22 nt, from 16 to 20 nt, from 16 to 18 nt, from 18 to 100 nt, from 18 to 80 nt, from 18 to 60 nt, from 18 to 40 nt, from 18 to 30 nt, from 18 to 25 nt, from 18 to 22 nt, or from 18 to 20 nt).

[0070] The guide sequence of a guide RNA can have a length of from 15 nt to 30 nt (e.g., 15 to 25 nt, 15 to 24 nt, 15 to 23 nt, 15 to 22 nt, 15 to 21 nt, 15 to 20 nt, 15 to 19 nt, 15 to 18 nt,17 to 30 nt, 17 to 25 nt, 17 to 24 nt, 17 to 23 nt, 17 to 22 nt, 17 to 21 nt, 17 to 20 nt, 17 to 19 nt, 17 to 18 nt, 18 to 30 nt, 18 to 25 nt, 18 to 24 nt, 18 to 23 nt, 18 to 22 nt, 18 to 21 nt, 18 to 20 nt, 18 to 19 nt, 19 to 30 nt, 19 to 25 nt, 19 to 24 nt, 19 to 23 nt, 19 to 22 nt, 19 to 21 nt, 19 to 20 nt, 20 to 30 nt, 20 to 25 nt, 20 to 24 nt, 20 to 23 nt, 20 to 22 nt, 20 to 21 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, 26 nt, 27 nt, 28 nt, 29 nt, or 30 nt). In some cases, the guide sequence has a length of 17 nt. In some cases, the guide sequence has a length of 18 nt. In some cases, the guide sequence has a length of 19 nt. In some cases, the guide sequence has a length of 20 nt. In some cases, the guide sequence has a length of 21 nt. In some cases, the guide sequence has a length of 22 nt. In some cases, the guide sequence has a length of 23 nt. In some cases, the guide sequence has a length of 24 nt.

[0071] In some cases, the guide sequence has a length of from 15 to 50 nucleotides (e.g., from 15 nucleotides (nt) to 20 nt, from 20 nt to 25 nt, from 25 nt to 30 nt, from 30 nt to 35 nt, from 35 nt to 40 nt, from 40 nt to 45 nt, or from 45 nt to 50 nt).

[0072] The protein-binding segment of a guide RNA can include two stretches of nucleotides that are complementary to one another and hybridize to form a double stranded RNA duplex (dsRNA duplex). Thus, in some cases, the protein-binding segment includes a dsRNA duplex.

[0073] In some cases, the dsRNA duplex region includes a range of from 5-25 base pairs (bp) (e.g., from 5-22, 5-20, 5-18, 5-15, 5-12, 5-10, 5-8, 8-25, 8-22, 8-18, 8-15, 8-12, 12-25, 12-22, 12-18, 12-15, 13-25, 13-22, 13-18, 13-15, 14-25, 14-22, 14-18, 14-15, 15-25, 15-22, 15-18, 17-25, 17-22, or 17-18 bp, e.g., 5 bp, 6 bp, 7 bp, 8 bp, 9 bp, 10 bp, etc.). In some cases, the dsRNA duplex region includes a range of from 6-15 base pairs (bp) (e.g., from 6-12, 6-10, or 6-8 bp, e.g., 6 bp, 7 bp, 8 bp, 9 bp, 10 bp, etc.). In some cases, the duplex region includes 5 or more bp (e.g., 6 or more, 7 or more, or 8 or more bp). In some cases, the duplex region includes 6 or more bp (e.g., 7 or more, or 8 or more bp). In some cases, not all nucleotides of the duplex region are paired, and therefore the duplex forming region can include a bulge. The term “bulge” herein is used to mean a stretch of nucleotides (which can be one nucleotide) that do not contribute to a double stranded duplex, but which are

surround 5' and 3' by nucleotides that do contribute, and as such a bulge is considered part of the duplex region. In some cases, the dsRNA includes 1 or more bulges (e.g., 2 or more, 3 or more, 4 or more bulges). In some cases, the dsRNA duplex includes 2 or more bulges (e.g., 3 or more, 4 or more bulges). In some cases, the dsRNA duplex includes 1-5 bulges (e.g., 1-4, 1-3, 2-5, 2-4, or 2-3 bulges).

[0074] A guide nucleic acid can include one or more of: i) a modified sugar; ii) a modified nucleobase; and iii) one or more non-natural internucleoside linkages.

[0075] Suitable non-natural internucleoside linkage include, e.g., a phosphorothioate, a phosphoramidate, a non-phosphodiester, a heteroatom, a chiral phosphorothioate, a phosphorodithioate, a phosphotriester, an aminoalkylphosphotriester, a 3'-alkylene phosphonates, a 5'-alkylene phosphonate, a chiral phosphonate, a phosphinate, a, a 3'-amino phosphoramidate, an aminoalkylphosphoramidate, a phosphorodiamidate, a thionophosphoramidate, a thionoalkylphosphonate, a thionoalkylphosphotriester, a selenophosphate, and a boranophosphate.

[0076] Suitable modified moieties include, e.g., locked nucleic acid (LNA) sugar moieties, 2'-substituted sugar moieties, 2'-O-methoxyethyl modified sugar moieties, one or more 2'-O-methyl modified sugar moieties, 2'-O-(2-methoxyethyl) modified sugar moieties, 2'-fluoro modified sugar moieties, 2'-dimethylaminooxyethoxy modified sugar moieties, and 2'-dimethylaminoethoxyethoxy modified sugar moieties.

[0077] Suitable modified nucleobases include, e.g., 5-methylcytosines; 5-hydroxymethyl cytosines; xanthines; hypoxanthines; 2-aminoadenines; 6-methyl derivatives of adenine; 6-methyl derivatives of guanine; 2-propyl derivatives of adenine; 2-propyl derivatives of guanine; 2-thiouracils; 2-thiothymines; 2-thiocytosines; 5-propynyl uracils; 5-propynyl cytosines; 6-azo uracils; 6-azo cytosines; 6-azo thymines; pseudouracils; 4-thiouracils; an 8-haloadenins; 8-aminoadenines; 8-thioladeninse; 8-thioalkyladenines; 8-hydroxyladenines; 8-haloguanines; 8-aminoguanines; 8-thiolguanines; 8-thioalkylguanines; 8-hydroxylguanines; 5-halouracils; 5-bromouracils; 5-trifluoromethyluracils; 5-halocytosines; 5-bromocytosines; 5-trifluoromethylcytosines; 5-substituted uracils; 5-substituted cytosines; 7-methylguanines; 7-methyladenines; 2-F-adenines; 2-amino-adenines; 8-azaguanines; 8-azaadenines; 7-deazaguanines; 7-deazaadenines; 3-deazaguanines; 3-deazaadenines; tricyclic pyrimidines; phenoxazine cytidines; phenothiazine cytidines; substituted phenoxazine cytidines; carbazole cytidines; pyridoindole cytidines; 7-deazaguanosines; 2-aminopyridines; 2-pyridones; 5-substituted pyrimidines; 6-azapyrimidines; N-2, N-6 or O-6 substituted purines; 2-aminopropyladenines; 5-propynyluracils; and 5-propynylcytosines.

[0078] In some embodiments, the engineered guide ribonucleic acid structure comprises at least one modification. In some embodiments, the at least one modification includes any of the modifications

disclosed herein, such as the nucleic acid modifications described herein. In some embodiments, the at least one modification includes 2'-O-methoxyethoxy (2'-OMe), 2'-O-(2-methoxyethyl) (2'-O-moe), 2'-fluoro (2'-F), locked nucleic acid (LNA), pseudouridine (ψ), phosphorothioate (PS) bond, phosphorodiamidate morpholino oligomer (PMO), and/or 2'-phosphorylation (2'-P). In some embodiments, the engineered guide ribonucleic acid structure comprises (i) a 5' end modification, (ii) a 3' end modification, or (iii) a 5' end modification and a 3' end modification.

[0079] In some embodiments, the engineered guide ribonucleic acid structure comprises a ribonucleic acid sequence comprising at least four hairpins comprising a stem and a loop. In some embodiments, the engineered guide ribonucleic acid structure comprises a sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the following non-degenerate nucleotide sequence: NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (**SEQ ID NO: 12**).

[0080] Guided by a CRISPR-Cas effector guide RNA, a CRISPR-Cas effector protein in some cases generates site-specific double strand breaks (DSBs) or single strand breaks (SSBs) (e.g., when the CRISPR-Cas effector protein is a nickase variant) within double-stranded DNA (dsDNA) target nucleic acids, which are repaired either by non-homologous end joining (NHEJ) or homology-directed recombination (HDR).

[0081] In some cases, a composition of the present disclosure comprises: i) a fusion polypeptide of the present disclosure; ii) a guide nucleic acid; and iii) a donor nucleic acid. A donor nucleic acid comprises a nucleotide sequence having homology to a target sequence of a target nucleic acid.

[0082] In some cases, contacting a target DNA (with a fusion polypeptide of the present disclosure and a CRISPR-Cas effector guide RNA) occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair. Thus, in some cases, a subject method includes contacting the target DNA with a donor polynucleotide (e.g., by introducing the donor polynucleotide into a cell), wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA. In some cases, the method does not comprise contacting a cell with a donor polynucleotide, and the target DNA is modified such that nucleotides within the target DNA are deleted.

[0083] In some cases, CRISPR-Cas effector guide RNA and a fusion polypeptide of the present disclosure are co-administered (e.g., contacted with a target nucleic acid, administered to cells, etc.) with a donor polynucleotide sequence that includes at least a segment with homology to the target DNA sequence, the subject methods may be used to add, i.e. insert or replace, nucleic acid material

to a target DNA sequence (e.g. to "knock in" a nucleic acid, e.g., one that encodes for a protein, an siRNA, an miRNA, etc.), to add a tag (e.g., 6xHis, a fluorescent protein (e.g., a green fluorescent protein; a yellow fluorescent protein, etc.), hemagglutinin (HA), FLAG, etc.), to add a regulatory sequence to a gene (e.g. promoter, polyadenylation signal, internal ribosome entry sequence (IRES), 2A peptide, start codon, stop codon, splice signal, localization signal, etc.), to modify a nucleic acid sequence (e.g., introduce a mutation, remove a disease causing mutation by introducing a correct sequence), and the like.

[0084] In applications in which it is desirable to insert a polynucleotide sequence into the genome where a target sequence is cleaved, a donor polynucleotide (a nucleic acid comprising a donor sequence) can also be provided to the cell. By a "donor sequence" or "donor polynucleotide" or "donor template" it is meant a nucleic acid sequence to be inserted at the site cleaved by the CRISPR-Cas effector protein (e.g., after dsDNA cleavage, after nicking a target DNA, after dual nicking a target DNA, and the like). The donor polynucleotide can contain sufficient homology to a genomic sequence at the target site, e.g. 70%, 80%, 85%, 90%, 95%, or 100% homology with the nucleotide sequences flanking the target site, e.g. within about 50 bases or less of the target site, e.g. within about 30 bases, within about 15 bases, within about 10 bases, within about 5 bases, or immediately flanking the target site, to support homology-directed repair between it and the genomic sequence to which it bears homology. Approximately 25, 50, 100, or 200 nucleotides, or more than 200 nucleotides, of sequence homology between a donor and a genomic sequence (or any integral value between 10 and 200 nucleotides, or more) can support homology-directed repair. Donor polynucleotides can be of any length, e.g. 10 nucleotides or more, 50 nucleotides or more, 100 nucleotides or more, 250 nucleotides or more, 500 nucleotides or more, 1000 nucleotides or more, 5000 nucleotides or more, etc.

[0085] The donor sequence is typically not identical to the genomic sequence that it replaces. Rather, the donor sequence may contain at least one or more single base changes, insertions, deletions, inversions or rearrangements with respect to the genomic sequence, so long as sufficient homology is present to support homology-directed repair (e.g., for gene correction, e.g., to convert a disease-causing base pair to a non-disease-causing base pair). In some embodiments, the donor sequence comprises a non-homologous sequence flanked by two regions of homology, such that homology-directed repair between the target DNA region and the two flanking sequences results in insertion of the non-homologous sequence at the target region. Donor sequences may also comprise a vector backbone containing sequences that are not homologous to the DNA region of interest and that are not intended for insertion into the DNA region of interest. Generally, the homologous region(s) of a donor sequence will have at least 50% sequence identity to a genomic sequence with

which recombination is desired. In certain embodiments, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or 99.9% sequence identity is present. Any value between 1% and 100% sequence identity can be present, depending upon the length of the donor polynucleotide.

[0086] The donor sequence may comprise certain sequence differences as compared to the genomic sequence, e.g. restriction sites, nucleotide polymorphisms, selectable markers (e.g., drug resistance genes, fluorescent proteins, enzymes etc.), etc., which may be used to assess for successful insertion of the donor sequence at the cleavage site or in some cases may be used for other purposes (e.g., to signify expression at the targeted genomic locus). In some cases, if located in a coding region, such nucleotide sequence differences will not change the amino acid sequence, or will make silent amino acid changes (i.e., changes which do not affect the structure or function of the protein). Alternatively, these sequences differences may include flanking recombination sequences such as FLPs, loxP sequences, or the like, that can be activated at a later time for removal of the marker sequence.

[0087] In some cases, the donor sequence is provided to the cell as single-stranded DNA. In some cases, the donor sequence is provided to the cell as double-stranded DNA. It may be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor sequence may be protected (e.g., from exonucleolytic degradation) by any convenient method and such methods are known to those of skill in the art. For example, one or more dideoxynucleotide residues can be added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides can be ligated to one or both ends. See, for example, Chang et al. (1987) Proc. Natl. Acad Sci USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues. As an alternative to protecting the termini of a linear donor sequence, additional lengths of sequence may be included outside of the regions of homology that can be degraded without impacting recombination. A donor sequence can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance. Moreover, donor sequences can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (e.g., adenovirus, AAV), as described elsewhere herein for nucleic acids encoding a CRISPR-Cas effector guide RNA and/or a CRISPR-Cas effector fusion polypeptide and/or donor polynucleotide.

[0088] A composition of the present disclosure can comprise, in addition to a fusion polypeptide of the present disclosure (and optionally also a CRISPR-Cas guide nucleic acid and optionally also a

donor nucleic acid), one or more of: a) a lipid; b) a buffer; c) a nuclease inhibitor; and d) a protease inhibitor.

[0089] The present disclosure provides a system comprising a fusion polypeptide of the present disclosure. A system of the present disclosure can comprise: a) a fusion polypeptide of the present disclosure and a CRISPR-Cas guide RNA; b) a fusion polypeptide of the present disclosure and a nucleic acid comprising a nucleotide sequence encoding a CRISPR-Cas guide RNA; c) a fusion polypeptide of the present disclosure, a CRISPR-Cas guide RNA, and a donor nucleic acid; or d) a fusion polypeptide of the present disclosure, a nucleic acid comprising a nucleotide sequence encoding a CRISPR-Cas guide RNA, and a donor nucleic acid.

Fusion polypeptides

[0090] In some aspects, provided herein are fusion polypeptides. In some embodiments, the fusion polypeptide comprises: i) an effector polypeptide (e.g., any effector polypeptide can be used in a fusion polypeptide of the present disclosure); and ii) one or more heterologous polypeptide (e.g., a targeting moiety that binds to an extracellular domain of a receptor); and optionally also includes a) an endosomal escape peptide (e.g., an EEP that can be used in a fusion polypeptide of the present disclosure); and/or b) one or more nuclear localization signals (NLSs). In some embodiments, the fusion polypeptide comprises: i) a CRISPR-Cas effector polypeptide (e.g., any CRISPR-Cas effector polypeptide can be used in a fusion polypeptide of the present disclosure); and ii) one or more heterologous polypeptide, e.g., a targeting moiety that binds to an extracellular domain of a receptor such as a transferrin protein 1 (TfR1, CD71)); and optionally also includes one or more nuclear localization signals (NLSs).

[0091] The term “heterologous polypeptide” is often used interchangeably herein with “fusion partner.” “Heterologous,” as used herein in the context of a polypeptide, can refer to an amino acid sequence that is not found in the native polypeptide. For example, a fusion CRISPR-Cas effector polypeptide comprises: a) a CRISPR-Cas effector polypeptide; and b) one or more heterologous polypeptides, where the heterologous polypeptide comprises an amino acid sequence from a protein other than a CRISPR-Cas effector polypeptide. “Heterologous,” as used herein in the context of a nucleic acid, refers to a nucleotide sequence that is not found in the native nucleic acid. As an example, in a guide nucleic acid, a heterologous guide nucleotide sequence (present in a targeting segment) that can hybridize with a target nucleotide sequence (target region) of a target nucleic acid is a nucleotide sequence that is not found in nature in a guide nucleic acid together with a binding segment that can bind to a CRISPR-Cas effector polypeptide of the present disclosure. For example, in some cases, a heterologous target nucleotide sequence (present in a heterologous targeting segment) is from a different source than a binding nucleotide sequence (present in a binding

segment) that can bind to a CRISPR-Cas effector polypeptide of the present disclosure. For example, a guide nucleic acid may comprise a guide nucleotide sequence (present in a targeting segment) that can hybridize with a target nucleotide sequence present in a eukaryotic target nucleic acid. A guide nucleic acid of the present disclosure can be generated by human intervention and can comprise a nucleotide sequence not found in a naturally-occurring guide nucleic acid. The term "naturally-occurring" as used herein as applied to a nucleic acid, a protein, a cell, or an organism, can refer to a nucleic acid, cell, protein, or organism that is found in nature.

[0092] In some cases, a fusion partner of an effector polypeptide provided herein can modulate transcription (e.g., inhibit transcription, increase transcription) of a target DNA. For example, in some cases the fusion partner is a protein (or a domain from a protein) that inhibits transcription (e.g., a transcriptional repressor, a protein that functions via recruitment of transcription inhibitor proteins, modification of target DNA such as methylation, recruitment of a DNA modifier, modulation of histones associated with target DNA, recruitment of a histone modifier such as those that modify acetylation and/or methylation of histones, and the like). In some cases, the fusion partner is a protein (or a domain from a protein) that increases transcription (e.g., a transcription activator, a protein that acts via recruitment of transcription activator proteins, modification of target DNA such as demethylation, recruitment of a DNA modifier, modulation of histones associated with target DNA, recruitment of a histone modifier such as those that modify acetylation and/or methylation of histones, and the like).

[0093] In some cases, a fusion partner of an effector polypeptide provided herein has enzymatic activity that modifies a target nucleic acid (e.g., nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity).

[0094] In some cases, a fusion partner of an effector polypeptide provided herein has enzymatic activity that modifies a polypeptide (e.g., a histone) associated with a target nucleic acid (e.g., methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity).

[0095] Examples of proteins (or fragments thereof) that can be used in increase transcription include but are not limited to: transcriptional activators such as VP16, VP64, VP48, VP160, p65 subdomain (e.g., from NFkB), and activation domain of EDLL and/or TAL activation domain (e.g., for activity

in plants); histone lysine methyltransferases such as SET1A, SET1B, MLL1 to 5, ASH1, SYMD2, NSD1, and the like; histone lysine demethylases such as JHDM2a/b, UTX, JMJD3, and the like; histone acetyltransferases such as GCN5, PCAF, CBP, p300, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, SRC1, ACTR, P160, CLOCK, and the like; and DNA demethylases such as Ten-Eleven Translocation (TET) dioxygenase 1 (TET1CD), TET1, DME, DML1, DML2, ROS1, and the like.

[0096] Examples of proteins (or fragments thereof) that can be used in decrease transcription include but are not limited to: transcriptional repressors such as the Krüppel associated box (KRAB or SKD); KOX1 repression domain; the Mad mSIN3 interaction domain (SID); the ERF repressor domain (ERD), the SRDX repression domain (e.g., for repression in plants), and the like; histone lysine methyltransferases such as Pr-SET7/8, SUV4-20H1, RIZ1, and the like; histone lysine demethylases such as JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, and the like; histone lysine deacetylases such as HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11, and the like; DNA methylases such as HhaI DNA m5c-methyltransferase (M.HhaI), DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3a (DNMT3a), DNA methyltransferase 3b (DNMT3b), METI, DRM3 (plants), ZMET2, CMT1, CMT2 (plants), and the like; and periphery recruitment elements such as Lamin A, Lamin B, and the like.

[0097] In some cases, a fusion partner of an effector polypeptide provided herein has enzymatic activity that modifies the target nucleic acid (e.g., ssRNA, dsRNA, ssDNA, dsDNA). Examples of enzymatic activity that can be provided by the fusion partner include but are not limited to: nuclease activity such as that provided by a restriction enzyme (e.g., FokI nuclease), methyltransferase activity such as that provided by a methyltransferase (e.g., HhaI DNA m5c-methyltransferase (M.HhaI), DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3a (DNMT3a), DNA methyltransferase 3b (DNMT3b), METI, DRM3 (plants), ZMET2, CMT1, CMT2 (plants), and the like); demethylase activity such as that provided by a demethylase (e.g., Ten-Eleven Translocation (TET) dioxygenase 1 (TET1CD), TET1, DME, DML1, DML2, ROS1, and the like), DNA repair activity, DNA damage activity, deamination activity such as that provided by a deaminase (e.g., a cytosine deaminase enzyme such as rat APOBEC1), dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity such as that provided by an integrase and/or resolvase (e.g., Gin invertase such as the hyperactive mutant of the Gin invertase, GinH106Y; human immunodeficiency virus type 1 integrase (IN); Tn3 resolvase; and the like), transposase activity, recombinase activity such as that provided by a recombinase (e.g.,

catalytic domain of Gin recombinase), polymerase activity, ligase activity, helicase activity, photolyase activity, and glycosylase activity).

[0098] In some cases, a fusion partner of an effector polypeptide provided herein has enzymatic activity that modifies a protein associated with the target nucleic acid (e.g., ssRNA, dsRNA, ssDNA, dsDNA) (e.g., a histone, an RNA binding protein, a DNA binding protein, and the like). Examples of enzymatic activity (that modifies a protein associated with a target nucleic acid) that can be provided by the fusion partner include but are not limited to: methyltransferase activity such as that provided by a histone methyltransferase (HMT) (e.g., suppressor of variegation 3-9 homolog 1 (SUV39H1, also known as KMT1A), euchromatic histone lysine methyltransferase 2 (G9A, also known as KMT1C and EHMT2), SUV39H2, ESET/SETDB1, and the like, SET1A, SET1B, MLL1 to 5, ASH1, SYMD2, NSD1, DOT1L, Pr-SET7/8, SUV4-20H1, EZH2, RIZ1), demethylase activity such as that provided by a histone demethylase (e.g., Lysine Demethylase 1A (KDM1A also known as LSD1), JHDM2a/b, JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, UTX, JMJD3, and the like), acetyltransferase activity such as that provided by a histone acetylase transferase (e.g., catalytic core/fragment of the human acetyltransferase p300, GCN5, PCAF, CBP, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, HBO1/MYST2, HMOF/MYST1, SRC1, ACTR, P160, CLOCK, and the like), deacetylase activity such as that provided by a histone deacetylase (e.g., HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11, and the like), kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, and demyristoylation activity.

[0099] Additional examples of suitable fusion partners are dihydrofolate reductase (DHFR) destabilization domain (e.g., to generate a chemically controllable fusion CRISPR-Cas effector protein), and a chloroplast transit peptide.

[00100] Additional suitable heterologous polypeptides include, but are not limited to, a polypeptide that directly and/or indirectly provides for increased transcription and/or translation of a target nucleic acid (e.g., a transcription activator or a fragment thereof, a protein or fragment thereof that recruits a transcription activator, a small molecule/drug-responsive transcription and/or translation regulator, a translation-regulating protein, etc.). Non-limiting examples of heterologous polypeptides to accomplish increased or decreased transcription include transcription activator and transcription repressor domains.

[00101] Non-limiting examples of heterologous polypeptides for use when targeting ssRNA target nucleic acids include (but are not limited to): splicing factors (e.g., RS domains); protein translation

components (e.g., translation initiation, elongation, and/or release factors such as eIF4G); RNA methylases; RNA editing enzymes (e.g., RNA deaminases, such as adenosine deaminase acting on RNA (ADAR), including A to I and/or C to U editing enzymes); helicases; RNA-binding proteins; and the like. It is understood that a heterologous polypeptide can include the entire protein or in some cases can include a fragment of the protein (e.g., a functional domain).

[00102] A fusion partner can be any domain capable of interacting with ssRNA (which, for the purposes of this disclosure, includes intramolecular and/or intermolecular secondary structures, e.g., double-stranded RNA duplexes such as hairpins, stem-loops, etc.), whether transiently or irreversibly, directly or indirectly, including but not limited to an effector domain selected from the group consisting of: Endonucleases (for example RNase III, the CRR22 DYW domain, Dicer, and PIN (PilT N-terminus) domains from proteins such as SMG5 and SMG6); proteins and protein domains responsible for stimulating RNA cleavage (for example CPSF, CstF, CFIm and CFIIm); Exonucleases (for example XRN-1 or Exonuclease T); Deadenylases (for example HNT3); proteins and protein domains responsible for nonsense mediated RNA decay (for example UPF1, UPF2, UPF3, UPF3b, RNP S1, Y14, DEK, REF2, and SRm160); proteins and protein domains responsible for stabilizing RNA (for example PABP); proteins and protein domains responsible for repressing translation (for example Ago2 and Ago4); proteins and protein domains responsible for stimulating translation (for example Staufen); proteins and protein domains responsible for (e.g., capable of) modulating translation (e.g., translation factors such as initiation factors, elongation factors, release factors, etc., e.g., eIF4G); proteins and protein domains responsible for polyadenylation of RNA (for example PAP1, GLD-2, and Star- PAP); proteins and protein domains responsible for polyuridinylation of RNA (for example CI D1 and terminal uridylate transferase); proteins and protein domains responsible for RNA localization (for example from IMP1, ZBP1, She2p, She3p, and Bicaudal-D); proteins and protein domains responsible for nuclear retention of RNA (for example Rrp6); proteins and protein domains responsible for nuclear export of RNA (for example TAP, NXF1, THO, TREX, REF, and Aly); proteins and protein domains responsible for repression of RNA splicing (for example PTB, Sam68, and hnRNP A1); proteins and protein domains responsible for stimulation of RNA splicing (for example Serine/Arginine-rich (SR) domains); proteins and protein domains responsible for reducing the efficiency of transcription (for example FUS (TLS)); and proteins and protein domains responsible for stimulating transcription (for example CDK7 and HIV Tat). Alternatively, the effector domain may be selected from the group consisting of: Endonucleases; proteins and protein domains capable of stimulating RNA cleavage; Exonucleases; Deadenylases; proteins and protein domains having nonsense mediated RNA decay activity; proteins and protein domains capable of stabilizing RNA; proteins and protein domains

capable of repressing translation; proteins and protein domains capable of stimulating translation; proteins and protein domains capable of modulating translation (e.g., translation factors such as initiation factors, elongation factors, release factors, etc., e.g., eIF4G); proteins and protein domains capable of polyadenylation of RNA; proteins and protein domains capable of polyuridinylation of RNA; proteins and protein domains having RNA localization activity; proteins and protein domains capable of nuclear retention of RNA; proteins and protein domains having RNA nuclear export activity; proteins and protein domains capable of repression of RNA splicing; proteins and protein domains capable of stimulation of RNA splicing; proteins and protein domains capable of reducing the efficiency of transcription; and proteins and protein domains capable of stimulating transcription. Another suitable heterologous polypeptide is a PUF RNA-binding domain, which is described in more detail in WO2012068627, which is hereby incorporated by reference in its entirety.

[00103] Some RNA splicing factors that can be used (in whole or as fragments thereof) as heterologous polypeptides for a fusion Cas12L polypeptide have modular organization, with separate sequence-specific RNA binding modules and splicing effector domains. For example, members of the Serine/ Arginine-rich (SR) protein family contain N-terminal RNA recognition motifs (RRMs) that bind to exonic splicing enhancers (ESEs) in pre-mRNAs and C-terminal RS domains that promote exon inclusion. As another example, the hnRNP protein hnRNP Al binds to exonic splicing silencers (ESSs) through its RRM domains and inhibits exon inclusion through a C-terminal Glycine-rich domain. Some splicing factors can regulate alternative use of splice site (ss) by binding to regulatory sequences between the two alternative sites. For example, ASF/SF2 can recognize ESEs and promote the use of intron proximal sites, whereas hnRNP Al can bind to ESSs and shift splicing towards the use of intron distal sites. One application for such factors is to generate ESFs that modulate alternative splicing of endogenous genes, particularly disease associated genes. For example, Bcl-x pre-mRNA produces two splicing isoforms with two alternative 5' splice sites to encode proteins of opposite functions. The long splicing isoform Bcl-xL is a potent apoptosis inhibitor expressed in long-lived postmitotic cells and is up-regulated in many cancer cells, protecting cells against apoptotic signals. The short isoform Bcl-xS is a pro-apoptotic isoform and expressed at high levels in cells with a high turnover rate (e.g., developing lymphocytes). The ratio of the two Bcl-x splicing isoforms is regulated by multiple cώ-elements that are located in either the core exon region or the exon extension region (i.e., between the two alternative 5' splice sites). For more examples, see WO2010075303, which is hereby incorporated by reference in its entirety.

[00104] Further suitable fusion partners include, but are not limited to, proteins (or fragments thereof) that are boundary elements (e.g., CTCF), proteins and fragments thereof that provide

periphery recruitment (e.g., Lamin A, Lamin B, etc.), protein docking elements (e.g., FKBP/FRB, Pil1/Aby1, etc.).

[00105] In some cases, the one or more heterologous polypeptides comprise a nuclease. Suitable nucleases include, but are not limited to, a homing nuclease polypeptide; a FokI polypeptide; a transcription activator-like effector nuclease (TALEN) polypeptide; a MegaTAL polypeptide; a meganuclease polypeptide; a zinc finger nuclease (ZFN); an ARCUS nuclease; and the like. The meganuclease can be engineered from an LADLIDADG homing endonuclease (LHE). A megaTAL polypeptide can comprise a TALE DNA binding domain and an engineered meganuclease. See, e.g., WO 2004/067736 (homing endonuclease); Urnov et al. (2005) Nature 435:646 (ZFN); Mussolino et al. (2011) Nucle. Acids Res. 39:9283 (TALE nuclease); Boissel et al. (2013) Nucl. Acids Res. 42:2591 (MegaTAL).

[00106] In some cases, the fusion partner (heterologous polypeptide) is a reverse transcriptase. In some cases, the fusion partner is a base editor. In some cases, the fusion partner (heterologous polypeptide) is a deaminase. In some cases, the one or more heterologous polypeptides comprise a reverse transcriptase polypeptide. In some cases, the CRISPR-Cas effector polypeptide is catalytically inactive. Suitable reverse transcriptases include, e.g., a murine leukemia virus reverse transcriptase; a Rous sarcoma virus reverse transcriptase; a human immunodeficiency virus type I reverse transcriptase; a Moloney murine leukemia virus reverse transcriptase; and the like.

[00107] In some cases, the one or more heterologous polypeptides comprise a base editor. Suitable base editors include, e.g., an adenosine deaminase; a cytidine deaminase (e.g., an activation-induced cytidine deaminase (AID)); APOBEC3G; and the like); and the like. A suitable adenosine deaminase is any enzyme that is capable of deaminating adenosine in DNA. In some cases, the deaminase is a TadA deaminase.

[00108] In some cases, the one or more heterologous polypeptides comprise a transcription factor. In some cases, the one or more heterologous polypeptides are capable of binding to a transcription factor. A transcription factor can include: i) a DNA binding domain; and ii) a transcription activator. A transcription factor can include: i) a DNA binding domain; and ii) a transcription repressor. Suitable transcription factors include polypeptides that include a transcription activator or a transcription repressor domain (e.g., the Kruppel associated box (KRAB or SKD); the Mad mSIN3 interaction domain (SID); the ERF repressor domain (ERD), etc.); zinc-finger-based artificial transcription factors (see, e.g., Sera (2009) Adv. Drug Deliv. 61:513); TALE-based artificial transcription factors (see, e.g., Liu et al. (2013) Nat. Rev. Genetics 14:781); and the like. In some cases, the transcription factor comprises a VP64 polypeptide (transcriptional activation). In some cases, the transcription factor comprises a Krüppel-associated box (KRAB) polypeptide

(transcriptional repression). In some cases, the transcription factor comprises a Mad mSIN3 interaction domain (SID) polypeptide (transcriptional repression). In some cases, the transcription factor comprises an ERF repressor domain (ERD) polypeptide (transcriptional repression). For example, in some cases, the transcription factor is a transcriptional activator, where the transcriptional activator is GAL4-VP16.

[00109] In some cases, the one or more heterologous polypeptides comprise a recombinase. Suitable recombinases include, e.g., a Cre recombinase; a Hin recombinase; a Tre recombinase; a FLP recombinase; and the like.

Nuclear localization signal

[00110] A fusion polypeptide of the present disclosure can include, in addition to the CRISPR-Cas effector polypeptide, a nuclear localization signal (NLS). In some cases, the one or more heterologous polypeptides comprise a nuclear localization signal (NLS). In general, NLS (or multiple NLSs) are of sufficient strength to drive accumulation of the fusion protein in a detectable amount in the nucleus of a eukaryotic cell. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the fusion protein such that location within a cell may be visualized. Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly.

[00111] In some cases, a fusion polypeptide of the present disclosure includes (is fused to) a nuclear localization signal (NLS) (e.g., in some cases 2 or more, 3 or more, 4 or more, or 5 or more NLSs). Thus, in some cases, a fusion polypeptide of the present disclosure includes one or more NLSs (e.g., 2 or more, 3 or more, 4 or more, or 5 or more NLSs). In some cases, one or more NLSs (2 or more, 3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) an N- terminus and/or a C-terminus of the fusion polypeptide. In some cases, one or more NLSs (2 or more, 3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the N-terminus. In some cases, one or more NLSs (2 or more, 3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the C-terminus. In some cases, one or more NLSs (3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) both the N-terminus and the C- terminus. In some cases, an NLS is positioned at the N-terminus and an NLS is positioned at the C- terminus.

[00112] In some cases, a fusion polypeptide of the present disclosure comprises a single NLS at an N-terminus of the CRISPR-Cas effector polypeptide present in the fusion polypeptide. In some cases, the fusion polypeptide of the present disclosure comprises a single NLS at a C-terminus of the

CRISPR- Cas effector polypeptide present in the fusion polypeptide. In some cases, a fusion polypeptide of the present disclosure comprises two NLSs at the N-terminus of the CRISPR-Cas effector polypeptide present in the fusion polypeptide. In some cases, a fusion polypeptide of the present disclosure comprises 2 NLSs at the C-terminus of the CRISPR-Cas effector polypeptide present in the fusion polypeptide. Non-limiting examples of NLSs include NLS sequences in **Table 2**.

Table 2. Exemplary NLS amino acid sequences

SEQ ID NO.	Amino acid sequence	NLS source
13	PKKKRKV	SV40 virus large T-antigen
14	KRPAATKKAGQAKKKK	nucleoplasmin (e.g., the nucleoplasmin bipartite)
15	PAAKRVKLD	c-myc
16	RQRRNELKRSP	c-myc
17	NQSSNFGPMKGGNFGGRSSGPYGGGGQY FAKPRNQGGY	hRNPA1 M9
18	RMRIZFKNKGKDTAELRRRRVEVSVELRK AKKDEQILKRRNV	IBB domain from importin-alpha
19	VSRKRPRP	myoma T protein
20	PPKKARED	myoma T protein
21	PQPKKKPL	human p53
22	SALIKKKKKMAP	mouse c-abl IV
23	DRLRR	influenza virus NS1
24	PKQKKRK	influenza virus NS1
25	RKLKKKIKKL	Hepatitis virus delta antigen
26	REKKKFLKRR	mouse Mx1 protein
27	KRKGDEVDGVDEVAKKKSKK	human poly(ADP-ribose) polymerase
28	RKCLQAGMNLEARKTKK	steroid hormone receptors (human) glucocorticoid
	K(K/R)X(K/R), where X is any amino acid	synthetic

[00113] In some embodiments, the NLS comprises an amino acid sequence having an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 13-28**. In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of **SEQ ID NOs: 13-28**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at

most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 13-28**. In some embodiments, the NLS comprises 100% sequence identity to any one of **SEQ ID NOs: 13-28**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 13-28.** In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 13-28.**

[00114] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 13**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 13**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 13**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 13**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 13**.

[00115] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 14**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 14**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 14**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 14**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 14**.

[00116] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 15**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 15**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 15**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 15**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 15**.

[00117] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 16**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 16**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 16**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 16**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 16**.

[00118] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 17**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 17**. In some

embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 17**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 17**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 17**.

[00119] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 18**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 18**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 18**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 18**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 18**.

[00120] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 19**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 19**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 19**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 19**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 19**.

[00121] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%,

at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 20**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 20**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 20**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 20**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 20**.

[00122] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 21**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 21**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 21**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 21**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 21**.

[00123] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 22**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 22**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 22**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 22**. In some

embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 22**.

[00124] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 23**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 23**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 23**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 23**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 23**.

[00125] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 24**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 24**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 24**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 24**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 24**.

[00126] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 25**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%,

at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 25**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 25**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 25**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 25**.

[00127] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 26**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 26**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 26**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 26**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 26**.

[00128] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 27**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 27**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 27**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 27**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 27**.

[00129] In some embodiments, the NLS comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to **SEQ ID NO: 28**. In some embodiments, the NLS comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to **SEQ ID NO: 28**. In some embodiments, the NLS comprises 100% sequence identity to **SEQ ID NO: 28**. In some embodiments, the NLS comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5 or more amino acid substitutions or mutations relative to **SEQ ID NO: 28**. In some embodiments, the NLS comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5 amino acid substitutions or mutations relative to **SEQ ID NO: 28**.

[00130] In some embodiments, the one or more NLSs comprise the amino acid sequence K(K/R)X(K/R), where X is any amino acid.

Targeting moiety

[00131] A fusion polypeptide of the present disclosure can include, in addition to the effector polypeptide, a targeting moiety. In some embodiments the fusion polypeptide comprises an effector polypeptide, and a targeting moiety (e.g., a CRISPR-Cas effector polypeptide covalently linked to a targeting moiety and coupled to an EEP, see for example **FIG. 1**). In some embodiments an N-terminus of the effector polypeptide is covalently linked to a C-terminus of the targeting moiety. In some embodiments a C-terminus of the effector polypeptide is covalently linked to an N-terminus of the targeting moiety. In some embodiments the fusion polypeptide comprises an effector polypeptide, a targeting moiety, and an EEP (e.g., a CRISPR-Cas effector polypeptide covalently linked to a targeting moiety and an EEP, see for example **FIG. 1**). In some embodiments an N-terminus of the effector polypeptide is linked to a C-terminus of the targeting moiety and a C-terminus of the effector polypeptide is linked to an N-terminus of the EEP. In some embodiments an N-terminus of the effector polypeptide is linked to a C-terminus of the EEP and a C-terminus of the effector polypeptide is linked to an N-terminus of the targeting moiety. In some embodiments an N-terminus of the effector polypeptide is linked to a C-terminus of the targeting moiety and an N-terminus of the targeting moiety is linked to a C-terminus of the EEP. In some embodiments a C-terminus of the effector polypeptide is linked to an N-terminus of the targeting moiety and a C-terminus of the targeting moiety is linked to an N-terminus of the EEP.

[00132] In some embodiments the fusion polypeptide comprises a CRISPR-Cas effector polypeptide, an endosomal escape polypeptide, and one or more heterologous polypeptides. In some cases, the fusion polypeptide comprises a CRISPR-Cas effector polypeptide and a targeting moiety. In some embodiments the fusion polypeptide comprises a CRISPR-Cas effector polypeptide and one or more heterologous polypeptides. In some embodiments at least one of the one or more heterologous polypeptides is a targeting moiety. A targeting moiety can be used to direct a polypeptide to a target cell or tissue. A targeting moiety in a fusion polypeptide disclosed herein can be but is not limited to a lipophilic moiety, a small molecule, a peptide (e.g., a polypeptide), an RNA molecule, a nanoparticle, an antibody, a single-domain antibody, a miniprotein, or an antigen binding fragment thereof. A targeting moiety in a fusion polypeptide disclosed herein can be specific to an antigen or receptor on the target cell or tissue (e.g., Asialoglycoprotein receptor (ASGPR)). In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein specifically binds to a target protein. In some embodiments, the target protein is a receptor expressed by a target cell. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to an extracellular domain of a receptor expressed by a target cell and facilitates cellular uptake. Non-limiting examples of a target protein can include IGF1R, IGF2R, TfR1, Nephrin, Neph1, Megalin, cKIT, GLUT4, GLUT1, Neuropilin, FOL1, FOL3, CCR5, CXCR1-4, EGFR, VEGFR1, VEGFR2, PDGFRA, ADRB3, NTRK2, GRIA1, GRIN1, GABRA1, and GABBR1.

[00133] A targeting moiety in a fusion polypeptide disclosed herein can be a lipophilic moiety. lipophilic moiety can comprise one or more fatty acid groups or salts thereof. A lipophilic be a lipid. Lipids are fatty acids and their derivatives which are insoluble in water but soluble in organic solvents. In some embodiments, a lipophilic moiety can be unsaturated. Alternatively, or in addition to, a lipophilic moiety can be monosaturated. Alternatively, or in addition to, a lipophilic moiety can be polysaturated. In some embodiments, a double bond of an unsaturated lipophilic moiety can be in a cis conformation. Alternatively, or in addition to, a double bond of an unsaturated lipophilic moiety can be in a trans conformation. Non-limiting examples of lipophilic moieties can be a triglyceride, a phospholipid, a sterol, an oil, a wax, a hormone, a vitamin, cholesterol, retinoic acid, cholic acid, adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-bis- 0(hexadecyl)glycerol, geranyloxyhexyanol, hexadecyl glycerol, bomeol, menthol, 1,3- propanediol, heptadecyl group, palmitic acid, myristic acid, 03-(oleoyl)lithocholic acid, 03- (oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine.

[00134] A targeting moiety in a fusion polypeptide disclosed herein can be a small molecule. A small molecule can be a sugar, an amino acid, a phenolic compound, an alkaloid, a sterol, a lipid, a fatty acid, or other small chemical compound. A chemical compound can be a molecule that has a

molecular weight of less than 1000 Daltons. Alternatively, or in addition to, a small molecule is a molecule with a size on the order of 1 nm.

[00135] A targeting moiety in a fusion polypeptide disclosed herein can be a sugar or sugar moiety. A sugar can be a monosaccharide. Alternatively, a sugar can be a disaccharide. Alternatively, a sugar can be a polysaccharide. Non-limiting examples of sugars include glucose, dextrose, fructose, galactose, a sugar alcohol, a pentose, xylose, ribose, sucrose, cellulose, starch, lactose, maltose, trehalose, lactulose, cellobiose, chitobiose, glycogen, or chitin. A small molecule can be an amino sugar such as but not limited to N-acetyl Galactosamine (GalNAc), N-acetylglucosamine or sialic acid.

[00136] A targeting moiety in a fusion polypeptide disclosed herein can be an RNA molecule. An RNA molecule can comprise an aptamer, a ribozyme, or a hairpin RNA.

[00137] A targeting moiety in a fusion polypeptide disclosed herein can be an antibody or an antigen-binding fragment thereof. An antibody, also known as an immunoglobulin, is a blood protein produced to counteract a specific antigen. Antibodies can be Y-shaped proteins which comprise variable binding sites that are specific to particular epitopes. An antibody can be a monoclonal antibody. Alternatively, an antibody can be a polyclonal antibody. An antibody can be a single-domain antibody. An antibody can be an antibody fragment. An antibody can be an agonist. Alternatively, an antibody can be an antagonist. Alternatively, an antibody can be an allosteric modulator (e.g., a positive allosteric modulator or a negative allosteric modulator).

[00138] In other cases, a targeting moiety disclosed herein is not an antibody or antigen-binding fragment thereof.

[00139] A targeting moiety in a fusion polypeptide disclosed herein can be a polypeptide. Non-limiting examples of polypeptides can be an agonist of IGF1R, IGF2R, TfR1, Nephrin, Neph1, Megalin, cKIT, GLUT4, GLUT1, Neuropilin, FOL1, FOL3, CCR5, CXCR1-4, EGFR, VEGFR1, VEGFR2, PDGFRA, or ADRB3.

[00140] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds a cell surface receptor. In certain embodiments, the cell surface receptor is membrane associated. Membrane proteins represent about a third of the proteins in living organisms and many membrane proteins are known in the field. Based on their structure, membrane proteins can be largely categorized into three main types: (1) integral membrane protein (IMP), which is permanently anchored or part of the membrane, (2) peripheral membrane protein, which is temporarily attached to the lipid bilayer or to other integral proteins, and (3) lipid-anchored proteins. The most common type of IMP is the transmembrane protein (TM), which spans the entire biological membrane. The cell surface receptor of the present disclosure includes single-pass and multi-pass membrane proteins.

Single-pass membrane proteins cross the membrane only once, while multi-pass membrane proteins weave in and out, crossing several times. In some embodiments, the cell surface receptor can be a monomeric receptor. In some embodiments, the cell surface receptor can be a multimeric receptor. In some embodiments, the cell surface receptor can form a complex with other molecules (e.g., an integrin). The cell surface receptor can be a recycling receptor. For example, a recycling receptor as used herein refers to a cell surface receptor that specifically binds to a ligand (e.g., a targeting moiety) and leads to internalization of the cell surface receptor.

[00141] In some case, the cell surface receptor can be selected by expression levels in one cell type relative to other cell types, where the cell surface receptor is enrichment in the selected cell type and/or tissue type. In some embodiments, the cell surface receptor comprises a tissue-type specific protein. In some embodiments, the cell surface receptor comprises a cell-type specific protein. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein is not an antibody or an antigen binding domain of an antibody. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein is an antibody or an antigen binding domain of an antibody.

[00142] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to an extracellular domain of a receptor protein. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to an extracellular domain of a receptor selected from the group consisting of: IGF1R, IGF2R, TfR1, Nephrin, Neph1, Megalin, cKIT, GLUT4, GLUT1, Neuropilin, FOL1, FOL3, CCR5, CXCR1-4, EGFR, VEGFR1, VEGFR2, PDGFRA, ADRB3, NTRK2, GRIA1, GRIN1, GABRA1, and GABBR1. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to an extracellular domain of transferrin receptor protein 1 (TfR1, CD71). In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to an extracellular domain of insulin-like growth factor 2 receptor (IGF2R). In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to an extracellular domain of insulin-like growth factor 1 receptor (IGF1R). In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises polypeptide sequence with less than 150 amino acids. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises polypeptide sequence with less than 100 amino acids.

[00143] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises polypeptide sequence with more than 50 amino acids. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises polypeptide sequence with more than 60 amino acids. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises polypeptide sequence with more than 70 amino acids. In some embodiments, targeting

moiety in a fusion polypeptide disclosed herein comprises polypeptide sequence with 80 amino acids or more.

[00144] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence with at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a polypeptide sequence selected from the sequences in **Table 3**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises 100% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 31-151** or **350-143854**.

[00145] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to TfR1. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises

an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 31-131** or **350-5272**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 31-131** or **350-5272**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 31-131** or **350-5272**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises 100% sequence identity to any one of **SEQ ID NOs: 31-131** or **350-5272**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence of any one of **SEQ ID NOs: 31-131** or **350-5272**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 31-131** or **350-5272**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 31-131** or **350-5272**.

[00146] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to IGF1R. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 132-145** or **5273-71413**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 132-**

145 or **5273-71413**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 132-145** or **5273-71413**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises 100% sequence identity to any one of **SEQ ID NOs: 132-145** or **5273-71413**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence of any one of **SEQ ID NOs: 132-145** or **5273-71413**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 132-145** or **5273-71413**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 132-145** or **5273-71413**.

[00147] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to IGF2R. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 146-151** or **71414-103440**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 146-151** or **71414-103440**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 146-151** or **71414-103440**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises 100% sequence identity to any one of **SEQ ID NOs: 146-151** or **71414-103440**. In some embodiments, the

targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence of any one of **SEQ ID NOs: 146-151** or **71414-103440**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 146-151** or **71414-103440**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 146-151** or **71414-103440**.

[00148] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to Dnmt3a. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 103441-111640**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 103441-111640**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 103441-111640**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises 100% sequence identity to any one of **SEQ ID NOs: 103441-111640**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence of any one of **SEQ ID NOs: 103441-111640**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 103441-111640**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence

that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 103441-111640**.

[00149] In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein binds to Nephrin. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 111641-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 111641-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 111641-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises 100% sequence identity to any one of **SEQ ID NOs: 111641-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises an amino acid sequence of any one of **SEQ ID NOs: 111641-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 111641-143854**. In some embodiments, the targeting moiety in a fusion polypeptide disclosed herein comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 111641-143854**.

[00150] Provided herein, in some aspects, is a suitable composition for target cell uptake and internalization of a CRISPR-Cas effector fusion polypeptide in the form of a ribonucleoprotein (RNP), which helps in an increase in cellular uptake of the fusion polypeptide (and thus of the

CRISPR-Cas effector polypeptide present in the fusion polypeptide) by at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100% (or two-fold), at least 2.5-fold, at least 5-fold, at least 10-fold, at least 25-fold, at least 50-fold, or more than 50-fold, compared to the cellular update of the same CRISPR-Cas effector polypeptide that not fused, or not fused the same configuration.

Table 3. Exemplary targeting moiety amino acid sequences

SEQ ID NO.	Amino acid sequence	Target protein
31	MVVYSDDPAVALEIATKYGGKPWEEIEKAAKEAAKKGVAVVVVVGTDEEETIAYNTALANDTGTGFEQAKKIADAIKAAVPNAKVVLA	TfR1
32	MAEEVYKLVEEMLEKGETDVKKIAKAVKEKFPEVNVIEVGNVAIFIVGGTEGLEERKAAAEKVYKEGIAAGWNTTELLTALLLATGADVVVA	TfR1
33	TADEVYKLVEEIGGSGPEAAERAAEAVEKKFPDVYVVRLPDGTAVFVVGHEGKEEKQAIALAVYEANKAAGITDPIERKINYLTAIPGAAVVYG	TfR1
34	AAAAAAARAAFEATLERLKADGTYDDPIAAGTALLLAVPGARVVVYDKKYAVFITDEKADEVYKLAEELLAKGATVEEISKAIKEAVPTVTVLLG	TfR1
35	GAALEEARRLAREVVERAKREGLDPLTTALRLLEAIPGASVVVAAESYAVFITKEKAEEVYKLAEEFDPKSAEDIRELAELVREKFPEVEVVLA	TfR1
36	GGEEAYELAEKLLAKGATPEEIAEAVAKELAGRVIVVRVGDVVVLVEGDEAEAPRRQALAEEVYERSRTLSPQERLEAFLTGVPGARVVVG	TfR1
37	AAARAAARAAFEGVLAQAKAEGITDPLEIALRLLEAIPDAAVVYANPYTVFVSREHADEVYRLVEKAIEEGITDSREIAKRVREAVPDVEVLLG	TfR1
38	EEERKELEEKVKKVLEEAKKEGITDPTEIAERILLETGARVVVAGEKYAAFFAPPEYAEEVYKLAEEALKKVKDVEEVARIVREKVPEVFVVLA	TfR1
39	AAEAAAKEAAARAVVEAGKKEGKSPLETAIALLEAVPGVRVVYYRERFAAFFAPEEDAPKVYKLVEEASKKYENEEEVAAYVKEKVPTVEVVIL	TfR1
40	AADEVYKLVDELLEKGETNVEVVAKAVREKFPEVTVITVGDVAVFIVGKEEEMEEREKKAKAVYEEAKKEGIDRETLLIRMLEATGADVVVG	TfR1
41	EEKWKEAEERFRKTLEKMKEEGVTDPAEIATALLLSIPDASVVVATDKYAVFIAPKEDHDKVFELAEKLLEEGGTDVADIAKKVQEAIPSVKVILG	TfR1
42	AADAVFAVAEELAGKASVEEIAEAIRERVEGNVLLIVKDDVLVAVVGEIADPEAVRARFDEVYEASKDLDPQSRLTAILLATGADVVVG	TfR1
43	LADKVYEIAESLLEFDDVKAIAEAVEKAVPGVHVVVIDDTAVFTVGGNKEEAEKKAKEVEEELKAKGITDKEERNLAFLEAIPNAKVVIS	TfR1
44	AAAREAAEAAVRAVLARAKAEGTSLEETNIALLEAVPDAAVVVGYERYAVFITKKDAEGVYKLAEENLPSSEEGIRELAERVREAFPDVHVIVD	TfR1
45	GMDKVYEIADKLIDEVENIEEIAKKIEEEVPGVRVIIIGNYAVAIEGGEENAEEYEAKAKKVIDEAKAKGITDPNEIGIALLLETGAGVVVS	TfR1
46	AADKVYAIAESSQGITDPREMAELVSAEVGGEVEVLVLDDFAVFVVGGSEDREEIEERLKEVVERGKKEGTPKEELAIELLLESGGAVVFT	TfR1
47	DAALEKAYKAASEALDKFGDNDMKGIAKYIADTVGDAVEVLYHKNVVYVKGGVVVSNNPAVALYLAENLAGAPKEEVFAAIDKLE	TfR1
48	MGEEVYKIAESALGKMSPEEVADLVEEKIGGKVRAIRLPGGYVVFVEGNDGFEEAEARARAAWEEIVAEGLTDDPTAVLTRLLLATGGSVAIA	TfR1
49	MEKEIIEEVKELLKTKTKEETAVALVEKFPEIKVVVVTDKIVAVVADDELGEKAYKAAEEALEKFPKDAAEEVAKYIKEKVGEGVKVISI	TfR1
50	MVVVSEDPELALKLVEELANGTKEEVEKFVKEYSGGGVKVVVVLTNDPETAEKAYKAAEEALDKFGINNTEEIAKYIKEKVGEDAEVVNI	TfR1
51	DELYEEALKLGEKIVEEGKKNGWSFEEIGERLLTAVKGAALVYVTEEKVVIVPEEVADEVYEAVSKMLEEGNFDEEEIIEKIKELAKGKAKVVVV	TfR1
52	GEKIEEFKKKFKAAMEENKNNPTTTPAEIALEAYPEAGLVIAYGDLVLLVHKEVADEAYKIAEEILGDDEESVVKAYEILKEKSGGKVTLVKI	TfR1

53	MAEKIYEIAEKNLEGHSVAEIGELIEKEVPGARTIVVGDVLAVVVGGTEDDATLRARAE AVVARGKAEGLSRLETLERILTEVPNAKVVVG	TfR1
54	EEAIEKLKEKIKEIVKEAKEKGLNTEETALRILTEVGAKLVVLTDGYVVIVLDPEDADKA YELAEELLKKYPNDSEIIVNKLKEAFPNVIVVEV	TfR1
55	ALDKVYEIAEKTLGKHNVKEVAEIVKKEVPEVEVIIVGDIAIFVVGGDPAEKRAAAEAV LERAKAEGLNRLETLLAFLEAVPGAGVVVG	TfR1
56	EELRKRVEAAVEEVWKQKENMNKEELATALLLKSGASVVVAAERYAVIVVPKEKAEE AYKIAEENLPENDEELEELAERIEEEVPGAVVVLT	TfR1
57	SAELAAKRAAFEAALAQSKSLNTTDALTLLLLATGASVGVWAGEYAVFLVPKEKEEEV YEIAEEMLKEGLKDVRELAERVEREVPGVVVLLG	TfR1
58	MEELEREERIVRELVEEGKREGIDKLTLAEKLLEAVPRARVVVVDEEYAVFIADPSIADK VYEIAEKALEEGVKDVREVAERVKKEVPGVHVIVA	TfR1
59	MELDEKAYEAAREALGTGENRGKKMAEYIKEKVPEVKVVLVGDDLVIVFRGDADVEA VVATAEKLLKEVSKEQAALELVEKFPEIIVVVV	TfR1
60	EEERERLRRLVEEVLERGKKEGVSPLELAEKLLLEIPGARVVAYDEKFAVFIADDETGEK VYKLAEEKLGKVSVEELAELVRKEFPSVEVLLA	TfR1
61	EEEFEKARAIFDAIVEEAKAKGWDAVTTGIALLEGVPGARVVAYTDRWAVFIADPKVA EEVYKLAEEAMEEVETVEEIAEVVKEKVPEVKVLLS	TfR1
62	DPLVEAALEAARIARERRAEDGGDDWVVVVIKDPEAAAEVAEAVGAEEVRQIGDYTVL FARGVSLEEAVAAAEAASRKYNSPYAVIGL	TfR1
63	MSEEQKAYKVAEKELFGDSEENAKRIGEAISKETGLKVIVHGKVVVVLKGEVDEAAVK AAIDALEGKSDEDRATALLQQFPEIVLVVV	TfR1
64	VHVIVIEPDENTVKIAEKVKEEAGNDPEEIYKVAEKLVDEYPTTGFAVFITDPTGAAEKE ATARAIYAEAKAKGVDRLTTLERLLLGTGAAVVVG	TfR1
65	GMDEELATQAVEIAAQVGAKKAYAIVGDGFILVIIVGKGAAAKKAAEKILELAKKAGIK GEVYMVEGGDDEELIEELLKKWEEIKAKLGK	TfR1
66	AAAQARADADVIVAHFGDLPERERTVLRLAAIAAAAGAEKVEAWEEGGRYHVRITGTD AAAKAGAAAARAWAAERGLDVEVTLG	TfR1
67	MGDKVYEVAEKLLEGSSPEKIAEAVKKAVPGVQVFVEGDFVAFVVGGEEDREEREARV REIVERAKREGWSTEETLLAILEGVPGATVVVG	TfR1
68	AYVVVVVSDGDLTPAAAAAAAEGATVRVEGNIVIAEVPDEATAMAVLTNAAAAVGDA RGFIIPAGEEGLEERIREEAERLRKREEEVRNGP	TfR1
69	MEAAEKAYEVAEKYFGTGEENAKKIAEAIKKEAGDELDVIVVKDIVIVGKGFDKEEVK KRVEELLKEKSLEETALELVEEFPKIYVVAV	TfR1
70	DAEFEAAEAAAKKVIEEAKKKGLIGKDPITVMTDLLLAIDGASTVVYDKEYAVFIHKEK ADEVYKLADELVGKYNVKEIAEIVKKEIPGVRVVML	TfR1
71	MVVVSDNPEAALAIAETNTGNMEEAKKMAKELTGGKGWIVVAFADDEKAEKAYKVA EKAIEIGDNKENAKKIAEAIRKEVPGVEVIEL	TfR1
72	MELEERVRKIVKEVVEKGKKEGKSPEEIAIAILEAVPEASVVYYSEDFVVLFPHKVAEEA YKLAEKLLEELTDEEEIKKKLLEEFPEAILVEL	TfR1
73	EKELEEKREKAKEIIEKGIKEGKSSVEIAEELLKEVGASVVYINDGVAVFLAPPEVADKA YKIVESFGGKKSAEEVAKAVVEETGGKVEVFLV	TfR1
74	AAAQARADADVIVAHFGDLPERERTVLRLAAIAAAAGAEKVEAWEEGGRYHVRITGTD EAAKEGAEAARAWAEERGLDVEVTLG	TfR1
75	EEKRKEAIEKIKEVIEKYKKTGDLEALGIAVIEKSGASVVVSLPNYVVAVLDPKLADEVY KLAEELIEKTDDEEEIAKAIKEAFPEVEVIKL	TfR1
76	EEERRAREARARELLAAARAAGIWESDPQALGTALLTGVPGARVVVVAERVAVFVTDE KADEVYELAEKNLPTSAEGVEELAELVRKELPDVEVILL	TfR1
77	LEEKVYELAEEALKTLTDVKEIAEYVKEKVGDKVTVFIIGDIAVFIVGDLSEEELEEKKK KAEEIYKQPLSTTDKMIEILKATGASVVVA	TfR1
78	MELLEKVYEVARKAIGKGENEGRVIAEAIREETGADVLLIGDGLVVVLTGNQDKEAIKE YYEKNKGDDEATAVEIAEKFGAGVVVV	TfR1
79	DEAAEKAYEAAEEALEKYPGREDGKKRADYIKEAVGDELDVRLFHPEVVVVGRGFDEE AVRAYVEKNLGDLEKLALELIEKEPKIVVVVV	TfR1
80	AARRVYEIVERAGGSGEEEAERVAEIVKKLVPDVVIFVHEDGFVVFVVNPEGAEEQRRR LEEVKEEAKRRGITDPQEIAVAYLEAVPGAAVVFR	TfR1
81	SELEQGVYELAEKLYGEGKENMERIAKAIKEKFPEFTVILVEDIVVVFRNADEEEIKEAV KRLLKEKSKEETALELVEKFPGVVVVVA	TfR1

82	DEEEIQKAIEELLKEKSLEEAAIEIAQRFGAAVVVVTKTLVVIKDDVILFGEGDREGKHIA EYIRRELEGFNGSKHERSTKAWEAAQKAF	TfR1
83	AARRVYEIVERANGGGAEEARRVADIVRKLVPEVEVLIVDPEFVVFVVGREGAEEQRRR LEEVRERARREGLSRQELAVARLEAIPKAAVVFG	TfR1
84	MADEVYRLAEEALGLTDVREKAAAVLARVGGRVRAFVFDDIAVFVDGGEDRWAEIRA RLEAVWRAIKARGGALTTEEKLTAFLEAVPDAVVAVA	TfR1
85	AVVVSENPEAALELVETVQDDPEKARALAKELTGGKKPIVVAFATEPELGQKVYDVAK EAFPNYGGEEGAKAIADAIKDTLGDEVTVIAF	TfR1
86	AAREVYEIVERALERGEEAERVAEIVKKLVGGEVTIFVHGDLVVFVVGREGEEEQRRRL EEVKEEAKRGLSPQEAAVLALEAVGAAVVYV	TfR1
87	MAAEEAAYKEAEKALEKGSGPELGRYAQEYLSKNVGGELEVHLGHPDIVVVHRGEDP AEIDAAFEELAHLSLEDRGIEIARRFPGVVVVVA	TfR1
88	VEVILIEPDEEIKKIAEEVKKEVPGGDPELIYKVAEKLTEKYPFKGTAVFVAGGLATVVS GGDTELATKLTLAAQGNAEQREANVKKVVEEYKK	TfR1
89	MVVVSEGDTETSIKLLLATQGDKESNEAAFKKVYAEAKAKGKGRFAVLVAPESVADEV YKVAEEYLNTEDLEEVAEKVKEAVPEVEVFLD	TfR1
90	SLFEALEKAAKLLKERSGSDDDVLVVIKNPEAAEDAAKAINAEEVTQIGDWTVLFARGV SFDEAVAAAEAASQKYNTPFAVLKL	TfR1
91	GVEEAYKIAEEALKGGNPKEIAKKIKEGVGDTIEVFDVGGEFAVVVPKGAGDVEARKA AAEKAYKAAKAAGETGAKAAIAIAEAVGVPSVVVA	TfR1
92	EEEFKEKEKIAKEVVEKGKKNGTPKEDILVDLLLATNARVVVSDPYVAFLADEENGEKV YEIAEKLVGKKNGKEIAEVVKKEVPNVKVIVL	TfR1
93	MEEIKKKEEKVEEVLERAKKEGITDPTEIALELLEATGASVVVYDEKYAVFIDKENAEEV YKIAEELLGGKTVEEIAEEVKKKIPNVKVFLK	TfR1
94	AAEKVYEIAERELGGGREDARRVAEIVRREVPDVVVVEVGDYVLFVVGRHADAERLRA IAEEALARVKAEGLENDPEAAGLLFLEAVPDAAVVVG	TfR1
95	GGDEAYKIALEAGENASVEEVAERVRRELEGRVIVLRFGDIAVFVDAEAGDEETHRERL EAAWEEALKIEDPVERAELFLLAVGASVVFR	TfR1
96	DVVVADDEEAALKLLEAAQGDEETIRKNVAEVAKELEGKGVKFVVFVADPEVADQVY EIADETIKKVKDPEEIAEAVKAKIGGKVLVIVE	TfR1
97	MVVVSFGDPQAALELVEKYAGAPLEEVKKAIKELGSEAIVYVYAEDDEAIEKAYEVAE KYFGGSGLEEAKRVAKKIEEEVGGKVEVVAE	TfR1
98	AARRVYEIVERANLGPEEAKRVADIVKKLVPDVIIFVHDGFVVFVVGREGEEEQRRRLE EVKEEAKRIGDPQELAVLYLEAVPGAAVVFR	TfR1
99	MVEEIKKAAKELSSLPAEEAAIELTEKFGLAVVVVDRLVIVTANSDELSTKAYEIAREVY GGGEEDSRRIAEAIKEALGPDVEVVIL	TfR1
100	DEREEEARRRVEEVKERAKREGLSRQELAVALLEAVPDAAVVVVTEEAVVFIVKDAAR RVYEIAERALPSSREELRRVADIVKKLVPDVIIIFH	TfR1
101	AAEEVYKIAEEVGGDKSAKEVAKAVKEAIGGKVYIIEFDDDTVVFFAGGGVVYGNNLE VNLKLLESIGKNKEETEKNVKETYEKIKN	TfR1
102	AAREVYEIVERANGGPAEEAERVAEIVKKLVPEVRIIRVAPDFVVFVVGDEGEEELRRRV EEVEERARREGWSLQERAVALLEAVPQAAVVFG	TfR1
103	MVVLTDNPEVGLELVEKLSQGTPEEVKAFIDEKLGGKGVVVYAVAETEEMQEKAYEV AEAAYRPLDAESAKEIAARIEEEVGEGVTVYSL	TfR1
104	EEEFKEAEKKAKEVIKKAKEEGVGETELLIRLLEAIPNARVVVANGYAVFLAEPEYADE VYKLAEKYGEKNNAEEVAEVVEKELKGKVRVIKV	TfR1
105	AAEKVYEIAEKEVVEDEEGIKELAEKVKEEVEGEVLVFVKGNIAVFVVGDFENKEEIEK KLEEAEKEIKEKGLSQTEALLTFLEKVNADVVYG	TfR1
106	MEMDERVYEVAEEALEASGGPEGAEAVARRVAEAVPGVKTIAVGTIVAIGRDFDEAAV RAAIEALAKTLSPEEAAEALIQAFPQITVVVV	TfR1
107	MVVFSVENVEFAEEVVENFSNKPMEEIEKYLKEKGAKDTVVAKATDPAKEEKAYEIAK KYFGSGPGEAEKIAEAIKEEVGEEDVLLV	TfR1
108	AARRVYEIVERARGEDAEEAKRVADIVKKLVPEVIIFVHDDTVVFVVGGNEEEQRRRLE EVKERAKREGWNRQDRAVLYLEAVPTAAVVYT	TfR1
109	MTEKVYELAEKEVGKDSIEEIAEAIKKAVPGVHVLLFDGKYAVFFSGGGVVVTEGDTEL AIKLLEAIGSTVEETEKNVKKVIEAAKK	TfR1
110	MVVVSMSGTEAALQLVEEFPNGDFEEIKKKAKELAKGQPLVVVRADDELGEKVYEAA REAFGGGDAEKAKEIGEYIKEKTGAEVIVL	TfR1

111	DALEKKVWEVASKALDLGSGPEAMKRIAAEIAAKTGAEVVAHGDIVYVSGGNVVVAL GGEEAALQLVEEVAGKPLEEVLAKMRELA	TfR1
112	MEEELKVYEVALKALEEAGENRGKQIAEYIKENAGDKFKVELISPDLVVVTDGDVDLE GIKKLYEEKKGDLEAAGVAIAEKFPGVVVVVV	TfR1
113	MVVVTDGDTEAALALLNAIGGTPEETRAAVEAALAEAKAKGVKAAVFHTGKNADEVY KLAEELLEVDDIEEIAEAVREKVPGVTVITL	TfR1
114	MARRVYEIVERARGKDEAKRVADIVKKLVPEVQIFVFGDVVVFVVGDPATEEEQRRRL EEVMEEGKREGLDKQTLAVRCLEAVNAAVVWV	TfR1
115	MVVVAVEGKEAALKLAEKYPNGDPEEIKKEAKKLTGGKPVVVAYGGTTNEEQLKIYK AAEEALEKNGDKEGAKEIAKYLKENGAKEVIVL	TfR1
116	MKEEVKKLAEELKDISEEEIAEELLRKFPEIVVVVANPLVVVFADDETGEKAYKAAEKAI GKGENEGENIAKYIKEEAGDDVEVILL	TfR1
117	MADEKVAVEAVEIAAQVGVKEAVVLVKDGKVYILVVGEGEKAKKALEEATKLAKELG FEVEARTIKPEELEEALEEGFRKL	TfR1
118	SEKARQAYEAARKNFKGGGTVVVLGGEVHAFGGGEEIKAQAEAIREESGGGGSVVVSD NPELALALVEKYTGDPEAVAKAVEELG	TfR1
119	MVVLTDDGADAAVELVANYTGDPAAVKAALKELTGGGITVVVVFAPDPELAEKVYKV AEKEAGKDSAEGMKEAAKKIKEETGATDVIVL	TfR1
120	HMEEIQKAIEELLKTISPEEAAIEIVQRFGAAVVVVTRDLVVVVCSDDHELSTRVWEAAQ KALETGGERQGKHIAEYIRREFEGEVEVILL	TfR1
121	MVVVVGGDTEEGEKFAKIMEEELGGVTTPEERAEKAYEIAVKHYHNAMVLVGVKDPE AVKKAWEELKGDNAKAAEELVEKFGIDVAVV	TfR1
122	MEEAEKAYKAAKEALEKFGKNKGKKIADYIKKNGGDKFKVIQLEPEIVVVHKGDEKEI KERVEELLKTLSPEEVGLKLAEEFGAVVVVV	TfR1
123	AADKVYEIAEKNLGSGEENAKRIAKLVEETLKGEVTVHMFGDVAVFVVGDPEKAAEVR ARLEAVEAELKAKGITDPVERRLAFLEAVGAAVVYG	TfR1
124	MEEEIQKAIEELLKTKSKEEAAIEIVQRFGTAVVVVSSELVVIHGGENHELSTRVWEAAQ KALETSGERDGKHIAEYIRRYTGGRVEVILI	TfR1
125	MEEEMKVYEAAREALEKAGENRGRHIAAYIREKAGDKFDVVLVGEDIVVAGKGLDKE AIKKEAEELRKIGLEEAAIKLAEKFGAIVVVV	TfR1
126	MVVVSTDPEAALTLVEKAGGGDMAEVDKLVAELTKGEPVVVVSGDEVRVFHPTPEAV AVAEAIRKAVGGGNEWEKAEAAWKAAEEAF	TfR1
127	MVVVSDNPEAALKIVENYQDDPEAVKKAVEELGVKKLVAVLNGEVHVFNPTPESKKIA DYVKEKIGGGDAEEKAEQAYKAAEEAW	TfR1
128	AADKVYELVESLLKEGVTDVNEIAEAVAKELPEVQVLIFGDVAVFVVGKEEKAEEIRAA AEKVWKEALEKGVGREERLIRLLLATGGAVAVG	TfR1
129	HETETRVWEAAQKALEETSGERQGKHISEYIRRYVPDVDVILHEGLVVISTGDVDEEEIK KAIEELLKKVSPEEAAIEIVQRFPEVSVVVV	TfR1
130	EEERERRERLVREVVERGKREGISKEELAIRVLEAIPDAAVVVATERIVAFITREDADKV YEIAEKNLGNSEEEDRRLAELVRREVPGVEVIFV	TfR1
131	MEAKEKAYEIAEEEYDKDSGPEAGKRIAEAIKKAIPGANVILIGERIVAVGVENEEEVRK KAEELEKEGLEKAAIELAEKFGIDVVVV	TfR1
132	GREEEIKELSELLNVPPEVAELMLKAEELLKKTGDPKVKQLLHYAASAAVQGYPELAEK LLKEAL	IGF1R
133	MDAIAEETAKRLGDPEAASKAVSVAYYLLQQGLAPEEVVEQLREIAKEFNDEAFAVVA EVLAELA	IGF1R
134	GKEEEIKEMAEMLGVPEEVARKLVELEELAKKTGDPRLKQLLHWAASAALTGNPEAM NAYLDEGL	IGF1R
135	SEKEEELIKFADYLESVYKGTGDPKYIEDLKELLEMAKELGAKKAVEYIEKKIKELGGSG GSGGS	IGF1R
136	DDAEFVYNAVKADPSILKSAIYALISQSKKEGNIEEVVKKLLELAKKAGDEAIAKAVKEL TKLVA	IGF1R
137	GAVALLKELVPLYEARGEPELAELVALALALLEAGNPSAAYAVLSYAAKESGDPRIQEV LDLIGA	IGF1R
138	GLREELEERLRETGDPEEVVREFIERLRKEFNISEKTALSAAWGALQALGLPELLEALAKI MEEV	IGF1R
139	DLLEEWKELLAKGVSPLEAAKLLIEKIAKEMNVSFKTALSALYGYASGEGNKELLEAIK ELMKEL	IGF1R

140	AEELAEDAKFILDSGKSLKFAFYVLLQELKNKGIPEEEAIPLVVEALAIASGDEELAKKV AEEVL	IGF1R
141	MEELEKLLKEAKKAIEEGNEKKANELLIAAWQKATKLGDEEKLKELEELIDLFAEKFGG SGGSGG	IGF1R
142	GAVKAAVDKALAGLNAELREDVTFILDHAVAEKNPGIALYFISRLAKETGDPALKALVA ELKKVA	IGF1R
143	GLQKAAQELLELIESSGNIYAAIGSYYGYARRKGDPALLTVVDTAAELVETDEAAALAY VRDIAA	IGF1R
144	AAERLARELRALGVPQDYVNVAVTLVATAARNPANTAAREALKVLIELLKEFNPEGAE VLARAAA	IGF1R
145	DEELIEEAERIAKEYQEKYGLSEKDFQVLFISAYQLLKKGVPEEEVRKIIEEMAEKLGGSG GSGG	IGF1R
146	DLEEKVEELVEEAEKTEDPEKRTALLNNAMLLAIRAKNEELIELVREARRELGGSGGSG GSGGSG	IGF2R
147	GEEAEEKVESLVEAAREEKDPTTRFNLLADAYIIAYKAKNPELVELVREARKELGGSGG SGGSGG	IGF2R
148	SDERIAEKLLRRARRLLEEGDEERAKLLLQLALMLAREAGTPLQEEIHRLARKLGGGSG GSGGSG	IGF2R
149	SEERVAEKLLRKAKELLEKGDPHRALLLLTSAKLFAKEAGDPELLREIHELARRIAGGSG GSGGS	IGF2R
150	DRSERSFERTRKRVEELEERGNPHEARLQLLSAFSVLRRLGDDELARKLQELLRRLIGGS GGSGG	IGF2R
151	GAAAREKVNALIEAALKEKDPDKASNLFNNAYILAYDAGDEAAIKAVREARRKKFGGS GGSGGSG	IGF2R

pH responsive polypeptide domain

[00151] A fusion polypeptide of the present disclosure can include, in addition to the effector polypeptide, a pH responsive polypeptide domain. In some embodiments the fusion polypeptide comprises an effector polypeptide, and a pH responsive polypeptide domain (e.g., a CRISPR-Cas effector polypeptide covalently linked to a pH responsive polypeptide domain and optionally coupled to an EEP). In some embodiments the fusion polypeptide comprises an effector polypeptide, a targeting moiety, and a pH responsive polypeptide domain (e.g., a CRISPR-Cas effector polypeptide covalently linked to a targeting moiety and a pH responsive polypeptide domain and optionally coupled to an EEP, see for example **FIG. 1**). In some embodiments an N-terminus of the effector polypeptide is linked to a C-terminus of the targeting moiety and a C-terminus of the effector polypeptide is linked to an N-terminus of the pH responsive polypeptide domain. In some embodiments an N-terminus of the effector polypeptide is linked to a C-terminus of the pH responsive polypeptide domain and a C-terminus of the effector polypeptide is linked to an N-terminus of the targeting moiety. In some embodiments an N-terminus of the effector polypeptide is linked to a C-terminus of the targeting moiety and an N-terminus of the targeting moiety is linked to a C-terminus of the pH responsive polypeptide domain. In some embodiments a C-terminus of the effector polypeptide is linked to an N-terminus of the targeting moiety and a C-terminus of the targeting moiety is linked to an N-terminus of the pH responsive polypeptide domain. In some embodiments the fusion polypeptide comprises a CRISPR-Cas effector polypeptide and one or more

heterologous polypeptides, where at least one of the one or more heterologous polypeptides is a pH responsive polypeptide domain.

[00152] In some embodiments, the pH responsive polypeptide domain is capable of functioning as an endosomal escape peptide as well, e.g., disrupting a membrane of a membrane-bound organelle (*e.g.*, an endosome). In some of these embodiments, the pH responsive polypeptide domain in a fusion polypeptide or polynucleotide disclosed herein is capable of disrupting a membrane of a membrane-bound organelle (*e.g.*, an endosome) and eliciting by itself endosomal escape of the fusion polypeptide or polynucleotide. In some embodiments, the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at a given pH level. The pH responsive polypeptide domain may bind to a EEP at neutral pH, but may be readily dissociated from it at acidic pH, for example, at pH 6.8, 6.7, 6.6, 6.5, 6.4, 6.3, 6.2, 6.1, 6.0, 5.9, 5.8, 5.7, 5.6, 5.5, 5.4, 5.3, 5.2, 5.1, 5.0 or lower. For example, at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4. In some embodiments, at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

[00153] In some embodiments, at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

[00154] In some embodiments, at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

[00155] In some embodiments, at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50,

60, 70, 80, 90, 100, 200, 500 or 1000 fold less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

[00156] In some embodiments, at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

[00157] In some embodiments, at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than the affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

[00158] In some embodiments, the endosomal escape peptide is released from the protein complex when the protein complex is present within an endosome.

[00159] In some embodiments, the endosomal escape peptide is released from the protein complex when the protein complex is present within a lysosome.

[00160] In some embodiments, the pH responsive polypeptide domain binds to or more endosomal escape peptides at pH 7.4 In some embodiments, the protein complex comprises two or more copies of the endosomal escape peptide.

[00161] In some embodiments, the pH responsive polypeptide domain comprises an amino acid sequence with at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a polypeptide sequence selected from the sequences in **Table 6**. In some embodiments, the pH responsive polypeptide domain comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 336-349**. In some embodiments, the pH responsive polypeptide domain comprises an amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 336-349**. In some embodiments, the pH responsive polypeptide domain comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 336-349**. In some embodiments, the pH responsive polypeptide domain comprises 100% sequence identity to any one of **SEQ ID NOs: 336-349**. In some embodiments, the pH responsive

polypeptide domain comprises a sequence to any one of **SEQ ID NOs: 336-349**. In some embodiments, the pH responsive polypeptide domain comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 336-349**. In some embodiments, the pH responsive polypeptide domain comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 336-349**.

[00162] In one aspect, provided herein is a composition comprising a fusion polypeptide comprising: (a) a CRISPR-Cas effector polypeptide; and (b) a targeting moiety that binds to an extracellular domain of a receptor expressed by a target cell, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody. In some embodiments, the fusion polypeptide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain binds to an endosomal escape peptide at pH 7.4. In some embodiments, the fusion polypeptide comprising: (a) a CRISPR-Cas effector polypeptide; and (b) a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain binds to an endosomal escape peptide (often referred to as an EEP binding peptide) at pH 7.4. In some embodiments, the fusion polypeptide further comprises a targeting moiety that binds to an extracellular domain of a receptor expressed by a target cell. In some embodiments, the targeting moiety is not an antibody or an antigen binding domain of an antibody. In some cases, a CRISPR-Cas effector polypeptide present in a fusion polypeptide of the present disclosure is itself a fusion polypeptide that comprises: i) a CRISPR-Cas effector polypeptide; and ii) a heterologous polypeptide, e.g., a targeting moiety binds to an extracellular domain of a receptor e.g., a transferrin protein 1 (TfR1, CD71) and a pH responsive polypeptide domain that can bind to an EEP in a pH responsive manner, and one or more NLS. Suitable CRISPR-Cas effector polypeptides include fusion polypeptides that comprise: i) a CRISPR-Cas effector polypeptide; and ii) a heterologous polypeptide, e.g., a targeting moiety binds to an extracellular domain of transferrin receptor protein 1 (TfR1, CD71) and a pH responsive polypeptide domain that can bind to an EEP in a pH responsive manner. Many CRISPR-Cas effector polypeptides are known to those skilled in the art, and any CRISPR-Cas effector polypeptide can be used in a fusion polypeptide of the present disclosure, such as any of the CRISPR-Cas effector polypeptides described herein.

Linker

[00163] In some cases, a fusion polypeptide of the present disclosure comprises a linker. In some embodiments, the linker comprises one or more linker polypeptides. The linker polypeptide may have any of a variety of amino acid sequences. Polypeptides of the present disclosure 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.

[00164] Examples of linker polypeptides include glycine polymers (G)n, glycine-serine polymers (including, for example, (GS)n, and (GGGGS)n, where n is an integer from 1 to 10), glycine-alanine polymers, serine polymers, and alanine-serine polymers. Exemplary linkers can comprise amino acid sequences including, but not limited to, GGS, GS, GGSG, GGSGG, GSGSG, GSGGG, GGGSG, GSSSG, and the like. 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. Suitable linkers include, e.g., GSSSSSSGS. A suitable spacer peptide includes, e.g., GIHGVPATT.

[00165] In some cases, a fusion polypeptide of the present disclosure comprises a linker between a targeting moiety and the polypeptide, such as a CRISPR-Cas effector polypeptide, present in the fusion polypeptide. In some cases, a fusion polynucleotide of the present disclosure comprises a linker between a targeting moiety and the polynucleotide, such as an siRNA present in the fusion polynucleotide. In some cases, the linker is a proteolytically cleavable linker. In some cases, a fusion polypeptide of the present disclosure comprises a first linker between a targeting moiety and a polypeptide disclosed herein, such as a CRISPR-Cas effector polypeptide, present in the fusion polypeptide, and a second linker between the polypeptide (or targeting moiety) and the compound disclosed herein. In some cases, a fusion polypeptide of the present disclosure comprises a first linker between a targeting moiety and a polynucleotide disclosed herein, such as an siRNA, present in the fusion polynucleotide, and a second linker between the polynucleotide (or targeting moiety)

and the compound disclosed herein. In some cases, the first and or second linker is a proteolytically cleavable linker.

[00166] The proteolytically cleavable linker can include a protease recognition sequence recognized by a protease selected from the group consisting of: alanine carboxypeptidase, Armillaria mellea astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, IgA-specific serine endopeptidase, legumain, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.

[00167] For example, the proteolytically cleavable linker can comprise a matrix metalloproteinase cleavage site, e.g., a cleavage site for a MMP selected from collagenase-1, -2, and -3 (MMP-1, -8, and -13), gelatinase A and B (MMP-2 and -9), stromelysin 1, 2, and 3 (MMP-3, -10, and -11), matrilysin (MMP-7), and membrane metalloproteinases (MT1-MMP and MT2-MMP). For example, the cleavage sequence of MMP-9 is Pro-X-X-Hy (wherein, X represents an arbitrary residue; Hy, a hydrophobic residue), e.g., Pro-X-X-Hy-(Ser/Thr), e.g., Pro-Leu/Gln-Gly-Met-Thr-Ser or Pro-Leu/Gln-Gly-Met-Thr. Another example of a protease cleavage site is a plasminogen activator cleavage site, e.g., a uPA or a tissue plasminogen activator (tPA) cleavage site. In some cases, the cleavage site is a furin cleavage site.

[00168] Specific examples of cleavage sequences of uPA and tPA include sequences comprising Val-Gly-Arg. Another example of a protease cleavage site that can be included in a proteolytically cleavable linker is a tobacco etch virus (TEV) protease cleavage site, e.g., ENLYTQS, where the protease cleaves between the glutamine and the serine. Another example of a protease cleavage site that can be included in a proteolytically cleavable linker is an enterokinase cleavage site, e.g., DDDDK, where cleavage occurs after the lysine residue. Another example of a protease cleavage site that can be included in a proteolytically cleavable linker is a thrombin cleavage site, e.g., LVPR. Additional suitable linkers comprising protease cleavage sites include linkers comprising one or more of the following amino acid sequences: LEVLFQGP, cleaved by PreScission protease (a fusion protein comprising human rhinovirus 3C protease and glutathione-S-transferase; Walker et al. (1994) Biotechnol. 12:601); a thrombin cleavage site, e.g., CGLVPAGSGP; SLLKSRMVPNFN or

SLLIARRMPNFN, cleaved by cathepsin B; SKLVQASASGVN or SSYLKASDAPDN, cleaved by an Epstein-Barr virus protease; RPKPQQFFGLMN cleaved by MMP-3 (stromelysin); SLRPLALWRSFN cleaved by MMP-7 (matrilysin); SPQGIAGQRNFN cleaved by MMP-9; DVDERDVRGFASFL cleaved by a thermolysin-like MMP; SLPLGLWAPNFN cleaved by matrix metalloproteinase 2(MMP-2); SLLIFRSWANFN cleaved by cathespin L; SGVVIATVIVIT cleaved by cathepsin D; SLGPQGIWGQFN cleaved by matrix metalloproteinase 1(MMP-1); KKSPGRVVGGSV cleaved by urokinase-type plasminogen activator; PQGLLGAPGILG cleaved by membrane type 1 matrixmetalloproteinase (MT-MMP); HGPEGLRVGFYESDVMGRGHARLVHVEEPHT cleaved by stromelysin 3 (or MMP-11), thermolysin, fibroblast collagenase and stromelysin-1; GPQGLAGQRGIV cleaved by matrix metalloproteinase 13 (collagenase-3); GGSGQRGRKALE cleaved by tissue-type plasminogen activator(tPA); SLSALLSSDIFN cleaved by human prostate-specific antigen; SLPRFKIIGGFN cleaved by kallikrein (hK3); SLLGIAVPGNFN cleaved by neutrophil elastase; and FFKNIVTPRTPP cleaved by calpain (calcium activated neutral protease).

[00169] In some embodiments, the linker is conjugated to the polypeptide and the heterologous polypeptide to produce the fusion polynucleotide. In some embodiments, conjugating comprises chemically conjugating. Many chemical approaches have been developed to crosslink carbohydrate and protein, all of which can be used to conjugate a compound disclosed herein to a polypeptide described herein, polynucleotide described herein, or any combination thereof. For example, the Staudinger ligation employs a substituted phosphite to react with the azide-modified protein to form the carbohydrate-protein conjugation via the amide bond formation (C. Grandjean, A. Boutonnier, C. Guerreiro, J. M. Fournier, L. A. Mulard, J Org Chem 2005, 70, 7123-7132). For example, oxime conjugation introduces an aminooxy group on the protein to react with oligosaccharide containing aldehyde or keto group (J. Kubler-Kielb, V. Pozsgay, J Org Chem 2005, 70, 6987-6990). For example, Michael addition often uses thiol group addition to maleimide to form a stable thioester linkage (T. Masuko, A. Minami, N. Iwasaki, T. Majima, S. Nishimura, Y. C. Lee, Biomacromolecules 2005, 6, 880-884). For example, the method of copper (I)-catalyzed cycloaddition of azide to alkyens (click chemistry) provides efficient glycoconjugation (a) H. C. Kolb, M. G. Finn, K. B. Sharpless, Angew Chem Int Ed Engl 2001, 40, 2004-2021; b) S. Hotha, S. Kashyap, J Org Chem 2006, 71, 364-367).

[00170] In some embodiments, chemically conjugating comprises a click chemistry reaction. Click chemistry comprises a group of biocompatible small molecule reactions commonly used in bioconjugation that are stereospecific, high yielding, wide in scope, and simple to perform. The primary click reaction utilizes copper-catalyzed coupling of a terminal alkyne with a terminal azide

to exclusively form the 1,2,3-triazole unit. These reactions create only byproducts that can be removed without chromatography, and can be conducted in easily removable or benign solvents. Examples of the click chemistry reaction includes, but are not limited to, Huisgen azide-alkyne 1,3-dipolar cycloaddition, copper-catalyzed azide-alkyne cycloaddition, ruthenium-catalyzed azide-alkyne cycloaddition, strain-promoted azide-alkyne cycloaddition, strain-promoted alkyne-nitrone cycloaddition, and reactions of strained alkenes such as alkene and azide [3+2] cycloaddition, alkene and tetrazine inverse-demand Diels-Alder, and alkene and tetrazole photoclick reaction. In some embodiments, the click chemistry reaction comprises a copper-catalyzed azide-alkyne cycloaddition. The mechanism of the copper-catalyzed azide-alkyne cycloaddition is described in Himo, et al. J. Am. Chem. Soc. 2005. 127: 210–216. In some embodiments, the click chemistry reaction to form a fusion polypeptide as disclosed herein occurs at ambient temperatures. In some embodiments, the click chemistry reaction to form a fusion polypeptide as disclosed herein occurs in the presence of a metal catalyst, for example, a copper(I)-catalyzed azide-alkyne cycloaddition. In some embodiments, the click chemistry reaction is performed in the absence of copper.

[00171] In some embodiments, compositions disclosed herein comprise a polypeptide (such as a CRISPR-Cas effector polypeptide as disclosed herein) and a heterologous polypeptide (e.g., to form a fusion polypeptide). In some cases, a fusion polypeptide of the present disclosure comprises a linker between a heterologous polypeptide (such as a targeting moiety) and the polypeptide, such as a CRISPR-Cas effector polypeptide, present in the fusion polypeptide. In some cases, the heterologous polypeptide (such as a targeting moiety disclosed herein) is conjugated to an N-terminus or a C-terminus of the polypeptide. In some cases, the heterologous polypeptide comprises a targeting moiety (such as a targeting moiety as disclosed herein). In some cases, the composition further comprises a pH responsive domain. In some cases, the polypeptide is conjugated to the targeting moiety or the pH responsive domain. In some cases, the polypeptide is conjugated to an N-terminus or a C-terminus of the targeting moiety. In some cases, the polypeptide is conjugated to an N-terminus or a C-terminus of the pH responsive domain.

[00172] The present disclosure provides fusion polypeptides comprising a CRISPR-Cas effector polypeptide and one or more heterologous polypeptides, where at least one of the one or more heterologous polypeptides is a targeting moiety that binds to an extracellular domain of a receptor expressed by a target cell and facilitates cellular uptake. The present disclosure also provides fusion polypeptides comprising a CRISPR-Cas effector polypeptide and one or more heterologous polypeptides, wherein at least one of the one or more heterologous polypeptides is a pH responsive polypeptide domain that binds to an endosomal escape peptide that facilitates endosomal escape of a complex comprising the fusion polypeptide. The present disclosure also provides fusion polypeptides

comprising a CRISPR-Cas effector polypeptide and two or more heterologous polypeptides, wherein at least one of the two or more heterologous polypeptides is a targeting moiety that binds to an extracellular domain of a receptor expressed by a target cell and facilitates cellular uptake, and at least one of the two or more heterologous polypeptides is a pH responsive polypeptide domain that binds to an endosomal escape peptide that facilitates endosomal escape of the CRISPR-Cas effector polypeptide.

[00173] The present disclosure provides compositions comprising a fusion polypeptide of the present disclosure and a guide nucleic acid. The present disclosure provides compositions and methods for delivery of RNPs in a cell. In one aspect, CRISPR-Cas-based RNPs are delivered using the compositions and methods described herein. In some embodiments, provided herein is an engineered RNP.

[00174] Provided herein, in some aspects, is a fused polypeptide, comprising a Cas endonuclease fused to a pH responsive binding peptide (EEP binder) that binds to an endosomal escape protein (EEP), such that the fused polypeptide is bound to an EEP at neutral pH, and such that when the fused polypeptide is in an endosomal compartment, the acidic pH of the endosomal compartment leads to the release of the EEP (and therefore the fused polypeptide) from the pH responsive binding peptide of the fused polypeptide. The released EEP can disrupt the endosome and therefore release the fused polypeptide from the endosome into the cytoplasm.

[00175] In some embodiments, the fusion polypeptide comprises a linker between the targeting moiety and the CRISPR-Cas effector polypeptide. In some embodiments, the linker between the targeting moiety and the EEP binder pH responsive polypeptide domain. In some embodiments, the linker between the targeting moiety and the CRISPR-Cas effector polypeptide is a proteolytically cleavable. In some embodiments, the fusion polypeptide comprises one or more nuclear localization sequences (NLSs). In some embodiments, the one or more NLSs comprise the amino acid sequence K(K/R)X(K/R), where X is any amino acid. In some embodiments, the one or more NLSs comprise the amino acid sequence PKKKRKV. In some embodiments, the one or more NLSs are at the N-terminus of the CRISPR-Cas effector polypeptide.

Polynucleotide

[00176] Disclosed herein in some embodiments are polymers comprising polynucleotides. The terms "polynucleotide" and "nucleic acid," can be used interchangeably herein, and can refer to a polymeric form of nucleotides of any length. In some embodiments a polymeric form of nucleotides can comprise ribonucleotides, deoxynucleotides, or combinations thereof. In some embodiments, a polynucleotide can comprise single-, double-, or multi-stranded DNA or RNA, genomic DNA,

cDNA, DNA-RNA hybrids, a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases, or any combination thereof. In some embodiments a polynucleotide can comprise single-stranded polynucleotides. In some embodiments a single-stranded polynucleotide can comprise a sense or antisense strand. In some embodiments, a polynucleotide can comprise a double-stranded polynucleotide. Unless specifically limited, a polynucleotide can comprise nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions), alleles, orthologs, SNPs, and complementary sequences as well as any sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); and Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)).

Nucleic acid modifications

[00177] In some embodiments, the polynucleotide can comprise any type of nucleic acid molecule (e.g., ribonucleic acid (RNA) molecule, deoxyribonucleic acid (DNA) molecule, xeno nucleic acid (XNA) molecule, etc.). In some embodiments, the polynucleotide comprises ribonucleic acid (RNA) molecules, deoxyribonucleic acid (DNA) molecules, xeno nucleic acid (XNA) molecules, or any combination thereof. In some embodiments, the polynucleotide is an engineered polynucleotide. An engineered polynucleotide can refer to a synthetically constructed polynucleotide. For example, the polynucleotide can be synthesized by chemical synthesis, enzymatic synthesis, or any combination thereof. In some embodiments, the polynucleotide is a synthetic polynucleotide. In some embodiments, the polynucleotide comprises at least one modification. In some embodiments, the polynucleotides disclosed herein can comprise nucleic acid molecules comprising a modification of a sugar, a phosphate backbone, or a nucleobase. For example, noncoding nucleic acid molecules can be modified through the addition of moieties onto the molecule. A modification can be a chemical modification, a synthetic modification, or a natural modification. For example, the polynucleotide disclosed herein can comprise a 2'-O-methoxyethoxy (2'-OMe) modification, a 2'-O-(2-methoxyethyl) (2'-O-moe) modification, a 2'-fluoro (2'-F) modification, a locked nucleic acid (LNA) modification, a pseudouridine (ψ) modification, a phosphorothioate (PS) bond modification, a phosphorodiamidate morpholino oligomer (PMO) modification, a 2'-phosphorylation (2'-P) modification, or any combination thereof. In some embodiments, the polynucleotide comprises (i) a

5' end modification, (ii) a 3' end modification, or (iii) a 5' end modification and a 3' end modification. In some embodiments, the modification is in the middle of the polynucleotide (e.g., between the 3' and the 5' end).

[00178] Nucleic acid molecules of the present disclosure can be modified at the nucleobase. Nucleobase modifications include but are not limited to 2'-O-methylation (2'-O-Me), conversion of uridine to pseudouridine, N(6)-methyladenosine, 5-methylcytidine, 5-methyluridine (ribothymidine), 2'-fluoro (2'F), 2'-O-methoxyethyl (2'-MOE), ribose modification with bridged nucleic acids (e.g., locked nucleic acids (LNA), ethylene-bridged nucleic acids (ENA), or constrained ethyl bridged nucleic acid (cEt) modifications), or nucleotides with alternative chemistries (e.g., phosphorodiamidate morpholino oligonucleotides (PMO), peptide nucleic acids (PNA), tricyclo DNA (tcDNA), unlocked nucleic acids (UNA) or glycol nucleic acids (GNA)).

[00179] Nucleic acid molecules of the present disclosure can be modified at the phosphate backbone. A phosphate backbone can be modified to comprise a phosphodiester, phosphorothioate isomers, phosphoryl DMI amidate diester isomers, a phosphorodithioate, a methylphosphontae, a 5'-phosphorothioate, a peptide nucleic acid, a 5'-(E)-vinylphosphonate, or a 5'-methyl phosphonate.

[00180] Additional moieties can be added or attached onto nucleic acid molecules of the present disclosure. Additional moieties can include, but are not limited to antibodies, lipophilic moieties, small molecules, and RNA aptamers (e.g., ribozymes, docosanoic acid, etc.). Additional moieties can be added to alter the pharmacological features (e.g., structural or chemical parameters) of the nucleic acid molecules of the present disclosure. Nucleic acid molecules of the present disclosure can be modified with 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 or more moieties. An additional moiety can be added at the 5' end of a nucleic acid molecule. Alternatively, an additional moiety can be added at the 3' end of a nucleic acid molecule. Alternatively, an additional moiety can be added in the middle of a nucleic acid molecule (e.g., neither at the 3' end nor the 5' end).

Effector nucleic acid molecules

[00181] In some embodiments, the polynucleotide is an effector nucleic acid molecule. "Effector nucleic acid molecules" can refer to nucleic acid sequences (e.g., microRNAs (miRNAs), small interfering RNAs (siRNAs), etc.) that directly interact with other molecules (e.g., RNA, polypeptides, etc.) to regulate gene expression (e.g., by inhibiting translation, causing degradation of a target mRNA, etc.), essentially acting as "effectors" to influence cellular processes. In some embodiments, the polynucleotide can be configured for polynucleotide-mediated gene expression alteration. Gene expression alteration can include, but is not limited to, changing the level of activity of a gene, such as turning it up (e.g., increasing expression) or down (e.g., decreasing expression)

through various mechanisms. Gene expression alterations can also refer to changes in gene regulation mechanisms, such as translation, transcription, epigenetics, post-transcription, post-translation, or any combination thereof. For example, in some cases, the polynucleotide can be configured for polynucleotide-mediated gene splicing, such as splice-switching oligonucleotides (SSOs) which can alter pre-mRNA splicing by binding to a target sequence and blocking the access of splicing factors to the pre-mRNA. Gene expression can be altered in a number of ways including, epigenetic alterations, DNA methylation, genome editing, site-directed mutagenesis, or any combination thereof. For example, in some cases, the polynucleotide can be configured for polynucleotide-mediated gene silencing, such as sequence-specific gene silencing. Sequence-specific gene silencing can refer to a decrease in gene expression relative to normal gene expression levels (e.g., in the absence of the polynucleotide disclosed herein). For example, polynucleotide-mediated gene silencing can refer to siRNA-mediated gene silencing (e.g., small interfering RNA, short interfering RNA, silencing RNA, etc.), where gene silencing can be elicited by binding of an siRNA, through complementarity, to a specific mRNA sequence, thereby targeting the specific mRNA sequence for cleavage and degradation.

[00182] In some embodiments, the polynucleotide can be complementary to a target nucleic acid sequence. In some embodiments, the polynucleotide can be complementary to a target ribonucleic acid sequence, such as a messenger ribonucleic acid sequence (mRNA), a pre-mRNA sequence, a coding ribonucleic acid sequence, a noncoding ribonucleic acid sequence, or any combination thereof. In some cases, the polynucleotide can selectively bind to a target amino acid sequence, such as an aptamer configured to bind a target protein. In some embodiments, the polynucleotide can comprise an siRNA, a microRNA (miRNA), an antisense oligonucleotide (ASO), a splice-switching oligonucleotide (SSO), an aptamer, a deoxyribozyme, or any combination thereof. The creation of such polynucleotides is routine to those of skill in the art, including strategies for developing and designing polynucleotides that will be effective for use in the methods disclosed herein (such as inhibition of a target gene). For example, factors such as melting temperature, length, GC percentage, secondary structure, target sequence, splicing enhancer sites, binding affinity, and modifications can be taken into account when designing polynucleotides of the present disclosure.

[00183] In some embodiments, the polynucleotide is an siRNA. Small interfering RNAs (siRNAs) generally comprise a noncoding double-stranded nucleic acid sequence that can be unwound into single stranded RNAs in order to bind to a target nucleic acid sequence, such as another RNA (e.g., mRNA) through Watson-Crick base pairing in a manner mediated by or independent of an argonaute-containing complex (e.g., RISC complex). An siRNA can function to specifically silence the expression of a target gene by binding to a complementary mRNA and causing degradation of the

complementary mRNA, thus preventing translation into a protein. siRNAs can be classified into classes I, II, and III based on predicted silencing activities. Class I siRNAs are functional in mammalian cells, while Class III siRNAs are nonfunctional. In some embodiments, the siRNA is a class I siRNA. Strategies for developing and designing siRNAs of the present disclosure that will be effective (such as in silencing the expression of a target gene) can include, but are not limited to, factors such as distance of target region to transcription start site (e.g., avoiding 5' and 3'UTRs), nucleotide composition (e.g., %GC content, length, etc.), and absence of off-target effects and secondary structures in a target site.

[00184] In some embodiments, the polynucleotide is a short hairpin RNA (shRNA). A short hairpin RNA (shRNA) can refer to a small, artificial RNA molecule that can silence genes by degrading viral or messenger RNA. shRNA is processed in the cytoplasm by the dicer protein, which removes a loop structure to produce siRNA. In some embodiments, the shRNA is configured to produce an siRNA, such as an siRNA disclosed herein.

[00185] In some embodiments, the polynucleotide is an miRNA. In some embodiments, the miRNA is a primary miRNA (pri-miRNA). In some embodiments, the miRNA is a precursor miRNA. In some embodiments, the miRNA is a mature miRNA. MicroRNA (miRNA) can refer to a small, non-coding RNA molecule that regulates gene expression by binding to a target mRNA and preventing or delaying the translation of the target mRNA. Strategies for developing and designing miRNAs of the present disclosure that will be effective (such as in silencing the expression of a target gene) can include, but are not limited to, factors such as seed regions (e.g., including at least one seed match in the 3′ UTR), secondary structure (e.g., including a terminal loop, a basal stem, and a flanking sequence on either side), and modifications.

[00186] In some embodiments, the polynucleotide can comprise an antisense oligonucleotide (ASO). Antisense oligonucleotides (ASOs) can refer to synthetic, single-stranded DNA or RNA molecules that bind to target RNA sequences and alter protein expression of the target RNA. For example, ASOs of the present disclosure can induce the degradation of mRNA and/or prevent or inhibit the progression of splicing or the translational machinery. ASOs can be modified to improve stability, uptake, and bioavailability. For example, modifications to the sugar ribose can increase affinity to the target RNA. Strategies for developing and designing ASOs of the present disclosure that will be effective (such as in inducing the degradation of a target gene) can include, but are not limited to, factors such as length, secondary structure, splicing enhancer sites, and modifications.

[00187] In some embodiments, the ASO is a splice-switching oligonucleotide (SSO). In some embodiments, the polynucleotide is an SSO. Splice-switching oligonucleotides (SSOs) are synthetic nucleic acids that can be used to disrupt the normal splicing process by blocking RNA-RNA base-

pairing and protein-RNA binding interactions. SSOs of the present disclosure can be designed to target different exons and have different lengths and modification contents. For example, AmNA (Amido-bridged nucleic acids) and GuNA (Guanidine-bridged nucleic acids) modified SSOs have been shown to have higher exon skipping activities than LNA (Locked nucleic acids) modified SSOs.

[00188] In some embodiments, the polynucleotide can comprise an aptamer. An aptamer can refer to a single-stranded nucleic acid molecule that binds to a target by folding into a three-dimensional structure. In some embodiments, the aptamer comprises DNA, RNA, XNA, or any combination thereof. In some embodiments, the aptamer binds to a polypeptide (e.g., protein), nucleic acid sequence, small molecule, or metal ion. In some embodiments, the aptamer is configured to alter gene expression. In some embodiments, the aptamer is an aptamer RNA switch. An RNA switch can refer to an RNA regulatory element that can change its structure (conformation) in response to the binding of a ligand (e.g., a polypeptide) in order to regulate gene expression of a target gene by either turning the target gene "on" or "off" depending on the ligand's presence. Strategies for developing and designing aptamers of the present disclosure that will be effective (such as in altering the expression of a target gene) can include, but are not limited to, factors such as secondary structure, and binding affinity.

[00189] In some embodiments, the polynucleotide is a deoxyribozyme (such as DNA enzymes, DNAzymes, catalytic DNA, etc.). Deoxyribozymes can refer to DNA sequences that can catalyze chemical reactions, including RNA and/or DNA cleavage, which can be used to control gene expression. In some embodiments, the deoxyribozyme catalyzes DNA cleavage. In some embodiments, the deoxyribozyme catalyzes RNA cleavage. In some embodiments, the deoxyribozyme is a ribonuclease.

Nucleic acid molecule structure

[00190] In some embodiments, the polynucleotide can comprise a double-stranded nucleic acid molecule (e.g., a double-stranded polynucleotide). In some embodiments, the double-stranded nucleic acid molecule is any type of nucleic acid (e.g., ribonucleic acid (RNA), deoxyribonucleic acid (DNA), etc.) that contains two strands. In some embodiments, the double-stranded nucleic acid molecules are double-stranded oligonucleotides, which comprise short pieces of a nucleotide sequence (e.g., DNA, RNA). In some embodiments, the double-stranded nucleic acid molecule comprises an antisense strand configured to silence a target single-stranded nucleic acid sequence, such as a messenger RNA (mRNA). In some embodiments, the double-stranded nucleic acid molecule comprises an antisense strand configured to enhance expression of a target nucleic acid sequence (such as a target nucleic acid sequence as described herein). In some embodiments, the

double-stranded nucleic acid molecule comprises an antisense strand configured to enhance expression of a single-stranded target nucleic acid sequence. The double-stranded nucleic acid molecule can be synthetic (e.g., engineered). Alternatively, or in addition to, the double-stranded nucleic acid molecule can be a chemically modified nucleic acid molecule, such as the modifications described herein. The double-stranded nucleic acid molecule can comprise two nucleic acid strands, a sense strand, and an antisense strand. The sense strand of the double-stranded nucleic acid molecule will have the same or substantially the same sequence as a target nucleic acid sequence. The antisense strand of the double-stranded nucleic acid molecule can be complementary or substantially complementary to a target nucleic acid sequence.

[00191] Double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be at least about 5 nucleotides, at least about 6 nucleotides, at least about 7 nucleotides, at least about 8 nucleotides, at least about 9 nucleotides, at least about 10 nucleotides, at least about 11 nucleotides, at least about 12 nucleotides, at least about 13 nucleotides, at least about 14 nucleotides, at least about 15 nucleotides, at least about 16 nucleotides, at least about 17 nucleotides, at least about 18 nucleotides, at least about 19 nucleotides, at least about 20 nucleotides, at least about 21 nucleotides, at least about 22 nucleotides, at least about 23 nucleotides, at least about 24 nucleotides, at least about 25 nucleotides, at least about 26 nucleotides, at least about 27 nucleotides, at least about 28 nucleotides, at least about 29 nucleotides, at least about 30 nucleotides, at least about 31 nucleotides, at least about 32 nucleotides, at least about 33 nucleotides, at least about 34 nucleotides, at least about 35 nucleotides, at least about 36 nucleotides, at least about 37 nucleotides, at least about 38 nucleotides, at least about 39 nucleotides, at least about 40 nucleotides, at least about 41 nucleotides, at least about 42 nucleotides, at least about 43 nucleotides, at least about 44 nucleotides, at least about 45 nucleotides, at least about 46 nucleotides, at least about 47 nucleotides, at least about 48 nucleotides, at least about 49 nucleotides, at least about 50 nucleotides, at least about 51 nucleotides, at least about 52 nucleotides, at least about 53 nucleotides, at least about 54 nucleotides, at least about 55 nucleotides, at least about 56 nucleotides, at least about 57 nucleotides, at least about 58 nucleotides, at least about 59 nucleotides, at least about 60 nucleotides, or more nucleotides in length.

[00192] Double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be at most about 60 nucleotides, at most about 59 nucleotides, at most about 58 nucleotides, at most about 57 nucleotides, at most about 56 nucleotides, at most about 55 nucleotides, at most about 54 nucleotides, at most about 53 nucleotides, at most about 52 nucleotides, at most about 51 nucleotides, at most about 50 nucleotides, at most about 49 nucleotides, at most about 48 nucleotides, at most about 47 nucleotides, at most about 46

nucleotides, at most about 45 nucleotides, at most about 44 nucleotides, at most about 43 nucleotides, at most about 42 nucleotides, at most about 41 nucleotides, at most about 40 nucleotides, at most about 39 nucleotides, at most about 38 nucleotides, at most about 37 nucleotides, at most about 36 nucleotides, at most about 35 nucleotides, at most about 34 nucleotides, at most about 33 nucleotides, at most about 32 nucleotides, at most about 31 nucleotides, at most about 30 nucleotides, at most about 29 nucleotides, at most about 28 nucleotides, at most about 27 nucleotides, at most about 26 nucleotides, at most about 25 nucleotides, at most about 24 nucleotides, at most about 23 nucleotides, at most about 22 nucleotides, at most about 21 nucleotides, at most about 20 nucleotides, at most about 19 nucleotides, at most about 18 nucleotides, at most about 17 nucleotides, at most about 16 nucleotides, at most about 15 nucleotides, at most about 14 nucleotides, at most about 13 nucleotides, at most about 12 nucleotides, at most about 11 nucleotides, at most about 10 nucleotides, at most about 9 nucleotides, at most about 8 nucleotides, at most about 7 nucleotides, at most about 6 nucleotides, at most about 5 nucleotides, or fewer nucleotides in length.

[00193] In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 55 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 50 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 45 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 40 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 35 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 30 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 25 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 20 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 15 nucleotides in

length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 5 to about 10 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 10 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 15 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 20 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 25 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 30 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 35 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 40 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 45 to about 60 nucleotides in length. In some embodiments, the double-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, etc.) can be between about 50 to about 60 nucleotides in length.

[00194] The antisense strand of a double-stranded noncoding nucleic acid molecule can form complementary binding with an entire sense strand of the double-stranded noncoding nucleic acid molecule. Alternatively, the antisense strand of a double-stranded noncoding nucleic acid molecule can form complementary binding with a part of a sense strand of a double-stranded noncoding nucleic acid molecule. The antisense strand of a double-stranded noncoding nucleic acid molecule can also form complementary binding with an entire target strand. Alternatively, the antisense strand of a double-stranded noncoding nucleic acid molecule can form complementary binding with a part of a target strand. The antisense strand of a double-stranded noncoding nucleic acid molecule can form complementary binding to one or more portions of a target strand (e.g., an mRNA strand) including a 5' UTR, a 3'UTR, a regulatory region, and/or a coding sequence. A regulatory region of a target strand (e.g., a mRNA) can comprise a promoter region, an enhancer region, an operator region, or a repressor region.

[00195] The antisense strand of the double-stranded noncoding nucleic acid molecule can form complementary binding to part or all of a portion of an mRNA target strand. The antisense strand of the double-stranded noncoding nucleic acid molecule can be completely complementary (e.g., 100% complementary) to their target strand counterparts. Alternatively, the antisense strand of the double-stranded noncoding nucleic acid molecule and can have imperfect complementarity to their target strand counterparts. An antisense strand of the double-stranded noncoding nucleic acid molecule can have at least about 50% complementarity, at least about 55% complementarity, at least about 60% complementarity, at least about 65% complementarity, at least about 70% complementarity, at least about 75% complementarity, at least about 80% complementarity, at least about 85% complementarity, at least about 90% complementarity, at least about 91% complementarity, at least about 92% complementarity, at least about 93% complementarity, at least about 94% complementarity, at least about 95% complementarity, at least about 96% complementarity, at least about 97% complementarity, at least about 98% complementarity, at least about 99% complementarity, or more to its corresponding target strand. An The antisense strand of the double-stranded noncoding nucleic acid molecule can have at most about 99% complementarity, at most about 98% complementarity, at most about 97% complementarity, at most about 96% complementarity, at most about 95% complementarity, at most about 94% complementarity, at most about 93% complementarity, at most about 92% complementarity, at most about 91% complementarity, at most about 90% complementarity, at most about 85% complementarity, at most about 80% complementarity, at most about 75% complementarity, at most about 70% complementarity, at most about 65% complementarity, at most about 60% complementarity, at most about 55% complementarity, at most about 50% complementarity, or less to its corresponding target strand.

[00196] In some embodiments, the polynucleotide can comprise a single-stranded nucleic acid molecule (e.g., a single-stranded polynucleotide). In some embodiments, the single-stranded nucleic acid molecule is any type of nucleic acid (e.g., ribonucleic acid (RNA), deoxyribonucleic acid (DNA), etc.) that contains one strand. In some embodiments, the single-stranded nucleic acid molecules are single-stranded oligonucleotides, which comprise short pieces of a nucleotide sequence (e.g., DNA, RNA). In some embodiments, the single-stranded nucleic acid molecule is configured to silence a target single-stranded nucleic acid sequence, such as a messenger RNA (mRNA). In some embodiments, the single-stranded nucleic acid molecule is configured to enhance expression of a target nucleic acid sequence (such as a target nucleic acid sequence as described herein). In some embodiments, the single-stranded nucleic acid molecule is configured to enhance expression of a single-stranded target nucleic acid sequence. The single-stranded nucleic acid

molecule can be synthetic (e.g., engineered). Alternatively, or in addition to, the single-stranded nucleic acid molecule can be a chemically modified nucleic acid molecule, such as the modifications described herein.

[00197] Single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be at least about 5 nucleotides, at least about 6 nucleotides, at least about 7 nucleotides, at least about 8 nucleotides, at least about 9 nucleotides, at least about 10 nucleotides, at least about 11 nucleotides, at least about 12 nucleotides, at least about 13 nucleotides, at least about 14 nucleotides, at least about 15 nucleotides, at least about 16 nucleotides, at least about 17 nucleotides, at least about 18 nucleotides, at least about 19 nucleotides, at least about 20 nucleotides, at least about 21 nucleotides, at least about 22 nucleotides, at least about 23 nucleotides, at least about 24 nucleotides, at least about 25 nucleotides, at least about 26 nucleotides, at least about 27 nucleotides, at least about 28 nucleotides, at least about 29 nucleotides, at least about 30 nucleotides, at least about 31 nucleotides, at least about 32 nucleotides, at least about 33 nucleotides, at least about 34 nucleotides, at least about 35 nucleotides, at least about 36 nucleotides, at least about 37 nucleotides, at least about 38 nucleotides, at least about 39 nucleotides, at least about 40 nucleotides, at least about 41 nucleotides, at least about 42 nucleotides, at least about 43 nucleotides, at least about 44 nucleotides, at least about 45 nucleotides, at least about 46 nucleotides, at least about 47 nucleotides, at least about 48 nucleotides, at least about 49 nucleotides, at least about 50 nucleotides, at least about 51 nucleotides, at least about 52 nucleotides, at least about 53 nucleotides, at least about 54 nucleotides, at least about 55 nucleotides, at least about 56 nucleotides, at least about 57 nucleotides, at least about 58 nucleotides, at least about 59 nucleotides, at least about 60 nucleotides, or more nucleotides in length.

[00198] Single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be at most about 60 nucleotides, at most about 59 nucleotides, at most about 58 nucleotides, at most about 57 nucleotides, at most about 56 nucleotides, at most about 55 nucleotides, at most about 54 nucleotides, at most about 53 nucleotides, at most about 52 nucleotides, at most about 51 nucleotides, at most about 50 nucleotides, at most about 49 nucleotides, at most about 48 nucleotides, at most about 47 nucleotides, at most about 46 nucleotides, at most about 45 nucleotides, at most about 44 nucleotides, at most about 43 nucleotides, at most about 42 nucleotides, at most about 41 nucleotides, at most about 40 nucleotides, at most about 39 nucleotides, at most about 38 nucleotides, at most about 37 nucleotides, at most about 36 nucleotides, at most about 35 nucleotides, at most about 34 nucleotides, at most about 33 nucleotides, at most about 32 nucleotides, at most about 31 nucleotides, at most about 30 nucleotides, at most about 29

nucleotides, at most about 28 nucleotides, at most about 27 nucleotides, at most about 26 nucleotides, at most about 25 nucleotides, at most about 24 nucleotides, at most about 23 nucleotides, at most about 22 nucleotides, at most about 21 nucleotides, at most about 20 nucleotides, at most about 19 nucleotides, at most about 18 nucleotides, at most about 17 nucleotides, at most about 16 nucleotides, at most about 15 nucleotides, at most about 14 nucleotides, at most about 13 nucleotides, at most about 12 nucleotides, at most about 11 nucleotides, at most about 10 nucleotides, at most about 9 nucleotides, at most about 8 nucleotides, at most about 7 nucleotides, at most about 6 nucleotides, at most about 5 nucleotides, or fewer nucleotides in length.

[00199] In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 55 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 50 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 45 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 40 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 35 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 30 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 25 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 20 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 15 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 5 to about 10 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs,

ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 10 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 15 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 20 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 25 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 30 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 35 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 40 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 45 to about 60 nucleotides in length. In some embodiments, the single-stranded nucleic acids of the present disclosure (e.g., siRNAs, miRNAs, ASOs, SSOs, aptamers, deoxyribozymes, etc.) can be between about 50 to about 60 nucleotides in length.

Fusion polynucleotides

[00200] In some aspects, the present disclosure relates to a fusion polynucleotide. Fusion polynucleotides can refer to polynucleotides that are fused with a protein, such as by fusing an endosomal escape peptide or a targeting moiety of the present disclosure to a polynucleotide of interest. In some embodiments, the fusion polynucleotide comprises: i) a polynucleotide (e.g., any polynucleotide disclosed herein can be used in a fusion polynucleotide of the present disclosure, such as an siRNA); and ii) one or more heterologous polypeptide, e.g., a targeting moiety that binds to an extracellular domain of a receptor such as a transferrin protein 1 (TfR1, CD71)); and optionally also includes one or more nuclear localization signals (NLSs), such as an NLS as disclosed herein (see for example NLS sequences in **Table 2**).

[00201] In some cases, nucleic acid modifications (such as nucleic acid modifications disclosed herein), and/or amino acid modifications (such as amino acid modifications disclosed herein) can be used to conjugate a polynucleotide (such as a polynucleotide disclosed herein) to a polypeptide (such as a targeting moiety disclosed herein). For example, the polynucleotide can be chemically modified with a reactive group at either the 5' or 3' end, then a bifunctional crosslinker can be used to attach

the polynucleotide to a reactive site on the polypeptide, such as a primary amine group on a lysine residue; common methods can include using click chemistry (e.g., azide-alkyne cycloaddition) or NHS ester coupling, where the polynucleotide is modified with an azide or amine group respectively, and the polypeptide is activated with a complementary reactive group like a DBCO (dibenzocyclooctyne) or NHS ester.

Targeting moiety

[00202] A fusion polynucleotide of the present disclosure can include, in addition to the polynucleotide, a targeting moiety. In some embodiments the fusion polynucleotide comprises a polynucleotide and a targeting moiety (e.g., an siRNA linked to a targeting moiety, see for example **FIG. 1**). In some embodiments a 5′ end of the polynucleotide is linked to a C-terminus of the targeting moiety. In some embodiments a 3′ end of the polynucleotide is linked to an N-terminus of the targeting moiety. In some embodiments the fusion polynucleotide comprises a polynucleotide, a targeting moiety, and an EEP (e.g., an siRNA linked to a targeting moiety and an EEP, see for example **FIG. 1**). In some embodiments a 5′ end of the polynucleotide is linked to a C-terminus of the targeting moiety and a 3′ end of the polynucleotide is linked to an N-terminus of the EEP. In some embodiments a 5′ end of the polynucleotide is linked to a C-terminus of the EEP and a 3′ end of the polynucleotide is linked to an N-terminus of the targeting moiety. In some embodiments a 5′ end of the polynucleotide is linked to a C-terminus of the targeting moiety and an N-terminus of the targeting moiety is linked to a C-terminus of the EEP. In some embodiments a 3′ end of the polynucleotide is linked to an N-terminus of the targeting moiety and a C-terminus of the targeting moiety is linked to an N-terminus of the EEP. In some embodiments the fusion polynucleotide comprises a polynucleotide, a targeting moiety, and a pH responsive polypeptide domain (e.g., an siRNA linked to a targeting moiety and a pH responsive polypeptide domain and coupled to an EEP, see for example **FIG. 1**). In some embodiments a 5′ end of the polynucleotide is linked to a C-terminus of the targeting moiety and a 3′ end of the polynucleotide is linked to an N-terminus of the pH responsive polypeptide domain. In some embodiments a 5′ end of the polynucleotide is linked to a C-terminus of the pH responsive polypeptide domain and a 3′ end of the polynucleotide is linked to an N-terminus of the targeting moiety. In some embodiments a 5′ end of the polynucleotide is linked to a C-terminus of the targeting moiety and an N-terminus of the targeting moiety is linked to a C-terminus of the pH responsive polypeptide domain. In some embodiments a 3′ end of the polynucleotide is linked to an N-terminus of the targeting moiety and a C-terminus of the targeting moiety is linked to an N-terminus of the pH responsive polypeptide domain.

[00203] In some embodiments the fusion polypeptide comprises a polynucleotide and one or more heterologous polypeptides, where at least one of the one or more heterologous polypeptides is a

targeting moiety. A targeting moiety can be used to direct a polynucleotide to a target cell or tissue. A targeting moiety in a fusion polynucleotide disclosed herein can be but is not limited to a lipophilic moiety, a small molecule, a peptide (e.g., a polypeptide), an RNA molecule, a nanoparticle, an antibody, a single-domain antibody, a miniprotein, or an antigen binding fragment thereof. A targeting moiety in a fusion polynucleotide disclosed herein can be specific to an antigen or receptor on the target cell or tissue (e.g., Asialoglycoprotein receptor (ASGPR)). In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein specifically binds to a target protein. In some embodiments, the target protein is a receptor expressed by a target cell. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein binds to an extracellular domain of a receptor expressed by a target cell and facilitates cellular uptake.

[00204] A targeting moiety in a fusion polynucleotide disclosed herein can be a lipophilic moiety. A lipophilic moiety can comprise one or more fatty acid groups or salts thereof. A lipophilic moiety can comprise one or more lipids. Lipids are fatty acids and their derivatives which are insoluble in water but soluble in organic solvents. In some embodiments, a lipophilic moiety can be unsaturated. Alternatively, or in addition to, a lipophilic moiety can be monosaturated. Alternatively, or in addition to, a lipophilic moiety can be polysaturated. In some embodiments, a double bond of an unsaturated lipophilic moiety can be in a cis conformation. Alternatively, or in addition to, a double bond of an unsaturated lipophilic moiety can be in a trans conformation. Non-limiting examples of lipophilic moieties can be a triglyceride, a phospholipid, a sterol, an oil, a wax, a hormone, a vitamin, cholesterol, retinoic acid, cholic acid, adamantane acetic acid, 1-pyrene butyric acid, dihydrotestosterone, 1,3-bis- 0(hexadecyl)glycerol, geranyloxyhexyanol, hexadecyl glycerol, bomeol, menthol, 1,3- propanediol, heptadecyl group, palmitic acid, myristic acid, 03-(oleoyl)lithocholic acid, 03- (oleoyl)cholenic acid, dimethoxytrityl, or phenoxazine.

[00205] A targeting moiety in a fusion polynucleotide disclosed herein can be a small molecule. A small molecule can be a sugar, an amino acid, a phenolic compound, an alkaloid, a sterol, a lipid, a fatty acid, or other small chemical compound. A chemical compound can be a molecule that has a molecular weight of less than 1000 Daltons. Alternatively, or in addition to, a small molecule is a molecule with a size on the order of 1 nm.

[00206] A targeting moiety in a fusion polynucleotide disclosed herein can be a sugar or sugar moiety. A sugar can be a monosaccharide. Alternatively, a sugar can be a disaccharide. Alternatively, a sugar can be a polysaccharide. Non-limiting examples of sugars include glucose, dextrose, fructose, galactose, a sugar alcohol, a pentose, xylose, ribose, sucrose, cellulose, starch, lactose, maltose, trehalose, lactulose, cellobiose, chitobiose, glycogen, or chitin. A small molecule

can be an amino sugar such as but not limited to N-acetyl Galactosamine (GalNAc), N-acetylglucosamine or sialic acid.

[00207] A targeting moiety in a fusion polynucleotide disclosed herein can be an RNA molecule. An RNA molecule can comprise an aptamer, a ribozyme, or a hairpin RNA.

[00208] A targeting moiety in a fusion polynucleotide disclosed hereincan be an antibody or an antigen-binding fragment thereof. An antibody, also known as an immunoglobulin, is a blood protein produced to counteract a specific antigen. Antibodies can be Y-shaped proteins which comprise variable binding sites that are specific to particular epitopes. An antibody can be a monoclonal antibody. Alternatively, an antibody can be a polyclonal antibody. An antibody can be a single-domain antibody. An antibody can be an antibody fragment. An antibody can be an agonist. Alternatively, an antibody can be an antagonist. Alternatively, an antibody can be an allosteric modulator (e.g., a positive allosteric modulator or a negative allosteric modulator).

[00209] A targeting moiety in a fusion polynucleotide disclosed herein can be a polypeptide. Non-limiting examples of polypeptides can be an agonist of IGF1R, IGF2R, TfR1, Nephrin, Neph1, Megalin, cKIT, GLUT4, GLUT1, Neuropilin, FOL1, FOL3, CCR5, CXCR1-4, EGFR, VEGFR1, VEGFR2, PDGFRA, or ADRB3.

[00210] In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein binds a cell surface receptor. In certain embodiments, the cell surface receptor is membrane associated. Membrane associated proteins represent about a third of the proteins in living organisms and many membrane proteins are known in the field. Based on their structure, membrane proteins can be largely categorized into three main types: (1) integral membrane protein (IMP), which is permanently anchored or part of the membrane, (2) peripheral membrane protein, which is temporarily attached to the lipid bilayer or to other integral proteins, and (3) lipid-anchored proteins. The most common type of IMP is the transmembrane protein (TM), which spans the entire biological membrane. The cell surface receptor of the present disclosure includes single-pass and multi-pass membrane proteins. Single-pass membrane proteins cross the membrane only once, while multi-pass membrane proteins weave in and out, crossing several times. In some embodiments, the cell surface receptor can be a monomeric receptor. In some embodiments, the cell surface receptor can be a multimeric receptor. In some embodiments, the cell surface receptor can form a complex with other molecules (e.g., an integrin). The cell surface receptor can be a recycling receptor. For example, a recycling receptor as used herein refers to a cell surface receptor that specifically binds to a ligand (e.g., a targeting moiety) and leads to internalization of the cell surface receptor.

[00211] In some case, the cell surface receptor can be selected by expression levels in one cell type relative to other cell types, where the cell surface receptor is enrichment in the selected cell type

and/or tissue type. In some embodiments, the cell surface receptor comprises a tissue-type specific protein. In some embodiments, the cell surface receptor comprises a cell-type specific protein. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein is not an antibody or an antigen binding domain of an antibody. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein is an antibody or an antigen binding domain of an antibody.

[00212] In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein binds to an extracellular domain of a receptor protein. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein binds to an extracellular domain of a receptor selected from the group consisting of: IGF1R, IGF2R, TfR1, Nephrin, Neph1, Megalin, cKIT, GLUT4, GLUT1, Neuropilin, FOL1, FOL3, CCR5, CXCR1-4, EGFR, VEGFR1, VEGFR2, PDGFRA, ADRB3, NTRK2, GRIA1, GRIN1, GABRA1, and GABBR1. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein binds to an extracellular domain of transferrin receptor protein 1 (TfR1, CD71). In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein binds to an extracellular domain of insulin-like growth factor 2 receptor (IGF2R). In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein binds to an extracellular domain of insulin-like growth factor 1 receptor (IGF1R).

[00213] In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises polypeptide sequence with less than 150 amino acids. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises polypeptide sequence with less than 100 amino acids. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises polypeptide sequence with more than 50 amino acids. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises polypeptide sequence with more than 60 amino acids. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises polypeptide sequence with more than 70 amino acids. In some embodiments, targeting moiety in a fusion polynucleotide disclosed herein comprises polypeptide sequence with 80 amino acids or more.

[00214] In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises an amino acid sequence with at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a polypeptide sequence selected from the sequences in **Table 3**. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises an

amino acid sequence having 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 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises an amino acid sequence having at most 40%, at most 45%, at most 50%, at most 55%, at most 60%, at most 65%, at most 70%, at most 75%, at most 80%, at most 85%, at most 90%, at most 91%, at most 92%, at most 93%, at most 94%, at most 95%, at most 96%, at most 97%, at most 98%, or at most 99% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises 100% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises a sequence to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises a sequence that has at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25 or more amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 31-151** or **350-143854**. In some embodiments, the targeting moiety in a fusion polynucleotide disclosed herein comprises a sequence that has at most 1, at most 2, at most 3, at most 4, at most 5, at most 10, at most 11, at most 12, at most 13, at most 14, at most 15, at most 16, at most 17, at most 18, at most 19, at most 20, at most 21, at most 22, at most 23, at most 24, or at most 25 amino acid substitutions or mutations relative to any one of **SEQ ID NOs: 31-151** or **350-143854**.

Linker

[00215] In some cases, a fusion polynucleotide of the present disclosure comprises a linker. In some embodiments, the linker is configured to link a polynucleotide of the present disclosure and a heterologous polypeptide of the present disclosure. In some cases, a fusion polynucleotide of the present disclosure comprises a linker between a heterologous polypeptide (such as a targeting moiety) and the polynucleotide, such as an siRNA, present in the fusion polynucleotide. In some cases, the linker is a proteolytically cleavable linker. In some cases, a polymer conjugate of the present disclosure comprises a linker between a heterologous polypeptide (such as a targeting moiety) and a polynucleotide disclosed herein, such as an siRNA, present in the fusion polynucleotide, and a linking group (as disclosed herein) between the polynucleotide (or heterologous polypeptide) and the compound disclosed herein. In some cases, the heterologous

polypeptide (such as a targeting moiety disclosed herein) is conjugated to a 5′ end or a 3′ end of the polynucleotide.

[00216] In some cases, the linker is configured to covalently link a polynucleotide of the present disclosure and a polypeptide of the present disclosure. In some cases, the polynucleotide, polypeptide, or a combination thereof comprise at least one reactive moiety. In some cases, the polynucleotide and the polypeptide are covalently linked by an enzymatic reaction. For example, SNAP-tag, a modified form of DNA repair enzyme named human O6alkylguanine-DNA-alkyltransferase, can be used to selectively form a covalent bond with a polynucleotide comprising a benzylguanine (e.g., RNA-TAG can be used to label the polynucleotide with a modified substrate analog, such as benzylguanine), thus allowing the conjugation of a polynucleotide of the present disclosure and a polypeptide of the present disclosure. As an additional example, RNAylation, mediated by the T4 phage ADP-ribosyltransferase (ART) ModB, can be used to covalently link the polynucleotide to the polypeptide through an N-glycosidic bond. In general, ARTs, a class of enzymes ubiquitous across all kingdoms of life, attach an ADP-ribose moiety from the redox cofactor nicotinamide adenine dinucleotide (NAD+, is referred as NAD in the following) to a target protein, nucleic acid, or small molecule in a covalent manner. ARTs generally exhibit high substrate specificity for NAD, such as a NAD-capped-RNA substrate. In some embodiments, a polynucleotide if the present disclosure comprises a NAD-cap (e.g., a NAD at a 5′ end of the polynucleotide).

[00217] In some cases, the linker is configured to noncovalently link a polynucleotide of the present disclosure and a polypeptide of the present disclosure. For example, a polypeptide of the present disclosure can comprise a targeting moiety and a nucleic acid (DNA, RNA, etc.) binding domain, wherein the nucleic acid binding domain is configured to bind a polynucleotide of the present disclosure (e.g., an siRNA, ASO, etc.). Thus, forming a noncovalent linkage between the polynucleotide and the polypeptide.

[00218] In some cases, a fusion polynucleotide of the present disclosure comprises a linker. In some embodiments, the linker is conjugated to the polynucleotide and the heterologous polypeptide to produce the fusion polynucleotide. In some embodiments, conjugating comprises chemically conjugating. Many chemical approaches have been developed to crosslink nucleotide and protein, all of which can be used to conjugate a polypeptide described herein, and a polynucleotide described herein. For example, oxime conjugation introduces an aminooxy group on the protein to react with oligosaccharide containing aldehyde or keto group (J. Kubler-Kielb, V. Pozsgay, J Org Chem 2005, 70, 6987-6990). For example, Michael addition often uses thiol group addition to maleimide to form a stable thioester linkage (T. Masuko, A. Minami, N. Iwasaki, T. Majima, S. Nishimura, Y. C. Lee, Biomacromolecules 2005, 6, 880-884). For example, the method of copper (I)-catalyzed

cycloaddition of azide to alkyens (click chemistry) provides efficient glycoconjugation (a) H. C. Kolb, M. G. Finn, K. B. Sharpless, Angew Chem Int Ed Engl 2001, 40, 2004-2021; b) S. Hotha, S. Kashyap, J Org Chem 2006, 71, 364-367).

[00219] In some embodiments, chemically conjugating comprises a click chemistry reaction. Click chemistry comprises a group of biocompatible small molecule reactions commonly used in bioconjugation that are stereospecific, high yielding, wide in scope, and simple to perform. The primary click reaction utilizes copper-catalyzed coupling of a terminal alkyne with a terminal azide to exclusively form the 1,2,3-triazole unit. These reactions create only byproducts that can be removed without chromatography, and can be conducted in easily removable or benign solvents. Examples of the click chemistry reaction includes, but are not limited to, Huisgen azide-alkyne 1,3-dipolar cycloaddition, copper-catalyzed azide-alkyne cycloaddition, ruthenium-catalyzed azide-alkyne cycloaddition, strain-promoted azide-alkyne cycloaddition, strain-promoted alkyne-nitrone cycloaddition, and reactions of strained alkenes such as alkene and azide [3+2] cycloaddition, alkene and tetrazine inverse-demand Diels-Alder, and alkene and tetrazole photoclick reaction. In some embodiments, the click chemistry reaction comprises a copper-catalyzed azide-alkyne cycloaddition. The mechanism of the copper-catalyzed azide-alkyne cycloaddition is described in Himo, et al. J. Am. Chem. Soc. 2005. 127: 210–216. In some embodiments, the click chemistry reaction to form a fusion polynucleotide as disclosed herein occurs at ambient temperatures. In some embodiments, the click chemistry reaction to form a fusion polynucleotide as disclosed herein occurs in the presence of a metal catalyst, for example, a copper(I)-catalyzed azide-alkyne cycloaddition. In some embodiments, the click chemistry reaction is performed in the absence of copper.

[00220] In some embodiments, compositions disclosed herein comprise a polynucleotide (such as a polynucleotide as disclosed herein) and a heterologous polypeptide (e.g., to form a fusion polynucleotide). In some cases, a fusion polynucleotide of the present disclosure comprises a linker between a heterologous polypeptide (such as a targeting moiety) and the polynucleotide, such as an siRNA, present in the fusion polynucleotide. In some cases, the heterologous polypeptide (such as a targeting moiety disclosed herein) is conjugated to a 5′ end or a 3′ end of the polynucleotide. In some cases, the heterologous polypeptide comprises a targeting moiety (such as a targeting moiety as disclosed herein). In some cases, the composition further comprises a pH responsive domain. In some cases, the polynucleotide is conjugated to the targeting moiety or the pH responsive domain. In some cases, the polynucleotide is conjugated to an N-terminus or a C-terminus of the targeting moiety. In some cases, the polynucleotide is conjugated to an N-terminus or a C-terminus of the pH responsive domain.

Nucleic acids, expression vectors, and modified host cells

[00221] The present disclosure provides a nucleic acid comprising a nucleotide sequence encoding a fusion polypeptide of the present disclosure. The present disclosure provides a nucleic acid comprising a polynucleotide of the present disclosure. The present disclosure provides a recombinant expression vector comprising a nucleic acid comprising a nucleotide sequence encoding a fusion polypeptide of the present disclosure. The present disclosure provides a recombinant expression vector comprising a polynucleotide of the present disclosure. The nucleic acids and recombinant expression vectors are useful for producing a fusion polypeptide of the present disclosure, a fusion polynucleotide of the present disclosure, or any combination thereof. In some cases, the nucleotide sequence encoding the fusion polypeptide, fusion polynucleotide, or any combination thereof is operably linked to one or more transcriptional control elements, e.g., promoters, such as promoters that are functional in a eukaryotic cell, where the promoter can be a constitutive promoter or an inducible promoter. "Operably linked" may refer to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression. As used herein, the terms "heterologous promoter" and "heterologous control regions" refer to promoters and other control regions that are not normally associated with a particular nucleic acid in nature. For example, a "transcriptional control region heterologous to a coding region" is a transcriptional control region that is not normally associated with the coding region in nature.

[00222] Suitable expression vectors are well known and include, but are not limited to, viral vectors (e.g., viral vectors based on vaccinia virus; poliovirus; adenovirus; adeno-associated virus; human immunodeficiency virus, and the like). Depending on the host/vector system utilized, any of a number of well-known, suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression. The terms "DNA regulatory sequences," "control elements," and "regulatory elements," may be used interchangeably herein, and can refer to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate expression of a coding sequence and/or production of an encoded polypeptide in a host cell.

[00223] The present disclosure provides a genetically modified host cell, where the host cell is genetically modified with a nucleic acid or a recombinant expression vector encoding a fusion polypeptide of the present disclosure. Suitable host cells include eukaryotic cells, such as yeast cells, insect cells, and mammalian cells. In some cases, the host cell is a cell of a mammalian cell line. Suitable mammalian cell lines include human cell lines, non-human primate cell lines, rodent (e.g.,

mouse, rat) cell lines, and the like. Suitable mammalian cell lines are likewise well known and include, but are not limited to, HeLa cells (e.g., American Type Culture Collection (ATCC) No. CCL-2), CHO cells (e.g., ATCC Nos. CRL9618, CCL61, CRL9096), 293 cells (e.g., ATCC No. CRL-1573), Vero cells, NIH 3T3 cells (e.g., ATCC No. CRL-1658), Huh-7 cells, BHK cells (e.g., ATCC No. CCL10), PC12 cells (ATCC No. CRL1721), COS cells, COS-7 cells (ATCC No. CRL1651), RAT1 cells, mouse L cells (ATCC No. CCLI.3), human embryonic kidney (HEK) cells (ATCC No. CRL1573), HLHepG2 cells, and the like.

[00224] A fusion polypeptide can be produced using a genetically modified host cell as described herein. Thus, this disclosure provides methods of producing a fusion polypeptide of the present disclosure. The methods generally involve culturing, in a culture medium, a host cell (an "expression host cell") that is genetically modified with a nucleic acid (e.g., a recombinant expression vector) comprising a nucleotide sequence encoding the fusion polypeptide; and isolating the fusion polypeptide from the genetically modified host cell and/or the culture medium.

[00225] Isolation of the fusion polypeptide from the expression host cell (e.g., from a lysate of the expression host cell) and/or the culture medium in which the host cell is cultured, can be carried out using standard methods of protein purification.

[00226] For example, a lysate may be prepared of the expression host and the lysate purified using high performance liquid chromatography (HPLC), exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. Alternatively, where the fusion polypeptide is secreted from the expression host cell into the culture medium, the fusion polypeptide can be purified from the culture medium using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. In some cases, the compositions which are used will comprise at least 80% by weight, at least about 85% by weight, at least about 95% by weight, or at least about 99.5% by weight, of the fusion polypeptide, in relation to contaminants related to the method of preparation of the product and its purification. The percentages can be based upon total protein.

[00227] In some cases, e.g., where the fusion polypeptide comprises an affinity tag, the fusion polypeptide can be purified using an immobilized binding partner of the affinity tag.

METHODS OF MODIFYING

[00228] The present disclosure provides methods of modifying a target nucleic acid and/or modifying a polypeptide associated with a target nucleic acid.

[00229] In some cases, the methods provided herein comprise editing a locus within a target cell, the method comprising introducing into the target cell one or more of the compositions disclosed herein. In some embodiments, the locus comprises two or more target nucleic acids (e.g., two or more target

sequences, two or more target genes, two or more target mRNAs, etc.). In some embodiments, the target cell is a eukaryotic cell. Suitable eukaryotic cells including mammalian cells, plant cells, insect cells, arachnid cells, protozoan cells, fish cells, fungal cells, yeast cells, amphibian cells, reptile cells, and avian cells. In some embodiments, the target cell is a tissue cell, such as cells from a kidney, brain, heart, or muscle.

[00230] In some embodiments, the target cell expresses a cell surface receptor. In some embodiments, the target cell expresses a cell-type specific receptor. In some embodiments, the target cell expresses a tissue-type specific receptor. In some embodiments, the targeting moiety disclosed herein binds the cell surface receptor. In some embodiments, the targeting moiety binds to an extracellular domain of the cell surface receptor. In some embodiments, the cell surface receptor is selected from the group consisting of: IGF1R, IGF2R, TfR1, Nephrin, Neph1, Megalin, cKIT, GLUT4, GLUT1, Neuropilin, FOL1, FOL3, CCR5, CXCR1-4, EGFR, VEGFR1, VEGFR2, PDGFRA, ADRB3, NTRK2, GRIA1, GRIN1, GABRA1, and GABBR1.

[00231] Provided herein, in some aspects, are methods of modifying a target nucleic acid (e.g., a target sequence, a target gene, etc.), the method comprising contacting the target nucleic acid with the one or more of the compositions disclosed herein. Also provided herein, in some aspects, are methods of modifying a target nucleic acid in a target cell, the method comprising introducing into the target cell one or more of the compositions disclosed herein. In some cases, a method of the present disclosure comprises contacting a eukaryotic cell comprising a target nucleic acid with a fusion polynucleotide of the present disclosure. In some cases, a method of the present disclosure comprises contacting a eukaryotic cell comprising a target nucleic acid with an RNP, where the RNP comprises: i) a fusion polypeptide of the present disclosure; and ii) a guide nucleic acid. In some cases, a method of the present disclosure comprises contacting a eukaryotic cell comprising a target nucleic acid with: a) an RNP, where the RNP comprises: i) a fusion polypeptide of the present disclosure; and ii) a guide nucleic acid; and b) a donor nucleic acid.

[00232] For example, a CRISPR-Cas effector fusion polypeptide of the present disclosure can be used to (i) modify (e.g., cleave, e.g., nick; methylate; deaminate; etc.) target nucleic acid (DNA or RNA; single stranded or double stranded); (ii) modulate transcription of a target nucleic acid; (iii) label a target nucleic acid; (iv) bind a target nucleic acid (e.g., for purposes of isolation, labeling, imaging, tracking, etc.); (v) modify a polypeptide (e.g., a histone) associated with a target nucleic acid; and the like. Thus, the present disclosure provides a method of modifying a target nucleic acid. In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting the target nucleic acid with: a) a CRISPR-Cas effector polypeptide of the present disclosure; and b) one or more (e.g., two) CRISPR-Cas effector guide RNAs. In some cases, a

method of the present disclosure for modifying a target nucleic acid comprises contacting the target nucleic acid with: a) a CRISPR-Cas effector polypeptide of the present disclosure; b) a CRISPR-Cas effector guide RNA; and c) a donor nucleic acid (e.g., a donor template). In some cases, the contacting step is carried out in a cell *in vitro*. In some cases, the contacting step is carried out in a cell *in vivo*. In some cases, the contacting step is carried out in a cell *ex vivo*. Modifications of a target nucleic acid that can be accomplished using a method of the present disclosure include, e.g., non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), and homology-directed repair (HDR). Such modifications can result in in gene knockout, DNA fragment insertion, deletion, replacement, or other modification. Modifications of a target nucleic acid that can be accomplished using a method of the present disclosure include base editing (e.g., modification of a cytidine or an adenosine). Modifications of a target nucleic acid that can be accomplished using a method of the present disclosure include any modification that can be carried out by a fusion partner, as described above (e.g., reverse transcription, base editing, etc.).

[00233] For example, the present disclosure provides (but is not limited to) methods of cleaving a target nucleic acid; methods of editing a target nucleic acid; methods of modulating transcription from a target nucleic acid; methods of isolating a target nucleic acid, methods of binding a target nucleic acid, methods of imaging a target nucleic acid, methods of modifying a target nucleic acid, and the like.

[00234] A polypeptide of the present disclosure, a polynucleotide of the present disclosure, a fusion polynucleotide of the present disclosure, or a fusion polypeptide of the present disclosure (e.g., a CRISPR-Cas effector fusion polypeptide when bound to a CRISPR-Cas effector guide RNA), can bind to a target nucleic acid, and in some cases, can bind to and modify a target nucleic acid. A target nucleic acid can be any nucleic acid (e.g., DNA, RNA), can be double stranded or single stranded, can be any type of nucleic acid (e.g., a chromosome (genomic DNA), derived from a chromosome, chromosomal DNA, plasmid, viral, extracellular, intracellular, mitochondrial, chloroplast, linear, circular, etc.) and can be from any organism (e.g., as long as the CRISPR-Cas effector guide RNA comprises a nucleotide sequence that hybridizes to a target sequence in a target nucleic acid, such that the target nucleic acid can be targeted).

[00235] A target nucleic acid can be DNA or RNA. A target nucleic acid can be double stranded (e.g., dsDNA, dsRNA) or single stranded (e.g., ssRNA, ssDNA). In some cases, a target nucleic acid is single stranded. In some cases, a target nucleic acid is a single stranded RNA (ssRNA). In some cases, a target ssRNA (e.g., a target cell ssRNA, a viral ssRNA, etc.) is selected from: mRNA, rRNA, tRNA, non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and microRNA (miRNA). In some cases, a target nucleic acid is a single stranded DNA (ssDNA) (e.g., a viral

DNA). As noted above, in some cases, a target nucleic acid is single stranded. A target nucleic acid can be genomic DNA (e.g., nuclear DNA). A target nucleic acid can be mitochondrial DNA. A target nucleic acid can be mitochondrial RNA. A target nucleic acid can be extrachromosomal DNA. A target nucleic acid sequence can be a coding RNA (e.g., an mRNA). Alternatively, a target nucleic acid sequence can be a non-coding RNA (e.g., a tRNA, an rRNA, an miRNA, a noncoding RNA, a small noncoding RNA, a long noncoding RNA, an siRNA, or a piwi-interacting RNA). Target nucleic acid sequences can be synthetic nucleic acid sequences. Alternatively, target nucleic acid sequence can be native to a cell or organism. In some embodiments, the target nucleic acid sequence is associated with a disease or a condition, such as a disease or a condition disclosed herein. For example, modulation of transcription of the target nucleic acid sequence (e.g., in the case of a DNA target) may be therapeutically effective to treat the disease or the condition. In another example, modulation of expression of the target nucleic acid sequence (e.g., in the case of an RNA target) may be therapeutically effective to treat the disease or the condition. In some embodiments, the modulation of the expression of the target nucleic acid sequence comprises post-transcriptional modifications, such as capping, splicing and polyadenylation of the RNA target.

[00236] In some embodiments, the target nucleic acid sequence comprises at least a portion of a target gene. In some embodiments, the target gene encodes a DNA repair protein (e.g., to inhibit unwanted microhomology-mediated end joining and/or nonhomologous end-joining and prevent proteins from functioning, or to recruit DNA repair enzymes to increase homology-directed repair), DNA methylation protein, histone methylation protein, histone demethylation protein, histone acetyltransferase (HAT) protein, histone deacetylation (HDAC) protein, histone crotonylation protein (e.g., an activating protein), histone phosphorylation protein, histone ubiquitination protein, or a receptor protein. In some embodiments, the target gene is selected from the group consisting of: 53BP1, LigIV, DNA-PK, CYREN, POLQ, Ku70, Ku80, XRCC4, RAD51, MRE11, RAD50, NBS1, CtIP (RBBP8), EXO1, DNA2, Dnmt1, UHRF1, UHFR2, Dnmt3a, Dnmt3b, Dnmt3L, MMSET, EZH2, SMYD3, SETDB1, NSD1, NSD3, SUV39H1/2, G9a/EHMT2, GLP, DOT1L, PRMT1-8, PRMT5, JHDM1/KDM2, LSD1/KDM1A, LSD2, KDM5B, HBO1, BRPF and JADE (e.g., subunits that determine which histone tails HBO1 acetylates), p300, CBP, GCN5, PCAF, SIRT1-7, HDAC1-11, Taf14, ATM Kinase, HASPIN, Rad6, Bre1, PRC1/BMI1, IGF1R, IGF2R, TfR1, Nephrin, Neph1, Megalin, cKIT, GLUT4, GLUT1, Neuropilin, FOL1, FOL3, CCR5, CXCR1-4, EGFR, VEGFR1, VEGFR2, PDGFRA, ADRB3, NTRK2, GRIA1, GRIN1, GABRA1, and GABBR1.

[00237] Modulation of a target nucleic acid can result in increased expression of a target gene, *e.g.*, expression of a protein encoded by the target gene, or expression of a functional RNA molecule (*e.g.*, transfer RNA (tRNA), ribosomal RNA (rRNA), microRNA, siRNA, piRNA, snoRNA, snRNA,

exRNA, scaRNA and the long non-codingRNA) encoded by the target gene. Alternatively, modulation of a target nucleic acid can result in decreased expression of the target gene. Alternatively, modulation of a target nucleic acid can maintain the expression of the target gene.

[00238] In some cases, methods disclosed herein can increase expression of a target gene by at least about 0.1%, at least about 0.2%, at least about 0.3%, at least about 0.4%, at least about 0.5%, at least about 0.6%, at least about 0.7%, at least about 0.8%, at least about 0.9%, at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, or more as compared to a control expression level.

[00239] In some cases, methods disclosed herein can increase expression of a target gene by at most about 500%, at most about 400%, at most about 300%, at most about 200%, at most about 100%, at most about 90%, at most about 80%, at most about 70%, at most about 60%, at most about 50%, at most about 40%, at most about 30%, at most about 20%, at most about 10%, at most about 9%, at most about 8%, at most about 7%, at most about 6%, at most about 5%, at most about 4%, at most about 3%, at most about 2%, at most about 1%, at most about 0.9%, at most about 0.8%, at most about 0.7%, at most about 0.6%, at most about 0.5%, at most about 0.4%, at most about 0.3%, at most about 0.2%, at most about 0.1%, or less as compared to a control expression level.

[00240] In some cases, methods disclosed herein can increase expression of a target gene by at least or up to about 0.1-fold, at least or up to about 0.2-fold, at least or up to about 0.3-fold, at least or up to about 0.4-fold, at least or up to about 0.5-fold, at least or up to about 0.6-fold, at least or up to about 0.7-fold, at least or up to about 0.8-fold, at least or up to about 0.9-fold, at least or up to about 1-fold, at least or up to about 2-fold, at least or up to about 3-fold, at least or up to about 4-fold, at least or up to about 5-fold, at least or up to about 6-fold, at least or up to about 7-fold, at least or up to about 8-fold, at least or up to about 9-fold, at least or up to about 10-fold, at least or up to about 20-fold, at least or up to about 30-fold, at least or up to about 40-fold, at least or up to about 50-fold, at least or up to about 60-fold, at least or up to about 70-fold, at least or up to about 80-fold, at least or up to about 90-fold, at least or up to about 100-fold, at least or up to about 500-fold, at least or up to about 1,000-fold, at least or up to about 5,000-fold, or at least or up to about 10,000-fold, as compared to a control expression level.

[00241] In some cases, methods disclosed herein can increase expression of a target gene by at most or less than about 10,000-fold, at most or less than about 5,000-fold, at most or less than about 1,000-fold, at most or less than about 500-fold, at most or less than about 100-fold, at most or less

than about 90-fold, at most or less than about 80-fold, at most or less than about 70-fold, at most or less than about 60-fold, at most or less than about 50-fold, at most or less than about 40-fold, at most or less than about 30-fold, at most or less than about 20-fold, at most or less than about 10-fold, at most or less than about 9-fold, at most or less than about 8-fold, at most or less than about 7-fold, at most or less than about 6-fold, at most or less than about 5-fold, at most or less than about 4-fold, at most or less than about 3-fold, at most or less than about 2-fold, at most or less than about 1-fold, at most or less than about 0.9-fold, at most or less than about 0.8-fold, at most or less than about 0.7-fold, at most or less than about 0.6-fold, at most or less than about 0.5-fold, at most or less than about 0.4-fold, at most or less than about 0.3-fold, at most or less than about 0.2-fold, at most or less than about 0.1-fold, as compared to a control expression level.

[00242] In some cases, methods disclosed herein can decrease expression of a target gene by at least about 0.1%, at least about 0.2%, at least about 0.3%, at least about 0.4%, at least about 0.5%, at least about 0.6%, at least about 0.7%, at least about 0.8%, at least about 0.9%, at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, or more as compared to a control expression level.

[00243] In some cases, methods disclosed herein can decrease expression of a target gene by at most about 500%, at most about 400%, at most about 300%, at most about 200%, at most about 100%, at most about 90%, at most about 80%, at most about 70%, at most about 60%, at most about 50%, at most about 40%, at most about 30%, at most about 20%, at most about 10%, at most about 9%, at most about 8%, at most about 7%, at most about 6%, at most about 5%, at most about 4%, at most about 3%, at most about 2%, at most about 1%, at most about 0.9%, at most about 0.8%, at most about 0.7%, at most about 0.6%, at most about 0.5%, at most about 0.4%, at most about 0.3%, at most about 0.2%, at most about 0.1%, or less as compared to a control expression level.

[00244] In some cases, methods disclosed herein can decrease expression of a target gene by at least or up to about 0.1-fold, at least or up to about 0.2-fold, at least or up to about 0.3-fold, at least or up to about 0.4-fold, at least or up to about 0.5-fold, at least or up to about 0.6-fold, at least or up to about 0.7-fold, at least or up to about 0.8-fold, at least or up to about 0.9-fold, at least or up to about 1-fold, at least or up to about 2-fold, at least or up to about 3-fold, at least or up to about 4-fold, at least or up to about 5-fold, at least or up to about 6-fold, at least or up to about 7-fold, at least or up to about 8-fold, at least or up to about 9-fold, at least or up to about 10-fold, at least or up to about 20-fold, at least or up to about 30-fold, at least or up to about 40-fold, at least or up to about 50-fold,

at least or up to about 60-fold, at least or up to about 70-fold, at least or up to about 80-fold, at least or up to about 90-fold, at least or up to about 100-fold, at least or up to about 500-fold, at least or up to about 1,000-fold, at least or up to about 5,000-fold, or at least or up to about 10,000-fold, as compared to a control expression level.

[00245] In some cases, methods disclosed herein can decrease expression of a target gene by at most or less than about 10,000-fold, at most or less than about 5,000-fold, at most or less than about 1,000-fold, at most or less than about 500-fold, at most or less than about 100-fold, at most or less than about 90-fold, at most or less than about 80-fold, at most or less than about 70-fold, at most or less than about 60-fold, at most or less than about 50-fold, at most or less than about 40-fold, at most or less than about 30-fold, at most or less than about 20-fold, at most or less than about 10-fold, at most or less than about 9-fold, at most or less than about 8-fold, at most or less than about 7-fold, at most or less than about 6-fold, at most or less than about 5-fold, at most or less than about 4-fold, at most or less than about 3-fold, at most or less than about 2-fold, at most or less than about 1-fold, at most or less than about 0.9-fold, at most or less than about 0.8-fold, at most or less than about 0.7-fold, at most or less than about 0.6-fold, at most or less than about 0.5-fold, at most or less than about 0.4-fold, at most or less than about 0.3-fold, at most or less than about 0.2-fold, at most or less than about 0.1-fold, as compared to a control expression level.

[00246] Modulation of a target nucleic acid can result in increased expression of a gene product expressed from a target mRNA. Alternatively, modulation of a target nucleic acid can result in decreased expression of the gene product expressed from the target mRNA. Alternatively, modulation of a target nucleic acid can maintain expression of the gene product expressed from the target mRNA.

[00247] In some cases, methods disclosed herein can increase expression of a target mRNA by at least about 0.1%, at least about 0.2%, at least about 0.3%, at least about 0.4%, at least about 0.5%, at least about 0.6%, at least about 0.7%, at least about 0.8%, at least about 0.9%, at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, or more as compared to a control expression level.

[00248] In some cases, methods disclosed herein can increase expression of a target mRNA by at most about 500%, at most about 400%, at most about 300%, at most about 200%, at most about 100%, at most about 90%, at most about 80%, at most about 70%, at most about 60%, at most about 50%, at most about 40%, at most about 30%, at most about 20%, at most about 10%, at most about

9%, at most about 8%, at most about 7%, at most about 6%, at most about 5%, at most about 4%, at most about 3%, at most about 2%, at most about 1%, at most about 0.9%, at most about 0.8%, at most about 0.7%, at most about 0.6%, at most about 0.5%, at most about 0.4%, at most about 0.3%, at most about 0.2%, at most about 0.1%, or less as compared to a control expression level.

[00249] In some cases, methods disclosed herein can increase expression of a target mRNA by at least or up to about 0.1-fold, at least or up to about 0.2-fold, at least or up to about 0.3-fold, at least or up to about 0.4-fold, at least or up to about 0.5-fold, at least or up to about 0.6-fold, at least or up to about 0.7-fold, at least or up to about 0.8-fold, at least or up to about 0.9-fold, at least or up to about 1-fold, at least or up to about 2-fold, at least or up to about 3-fold, at least or up to about 4-fold, at least or up to about 5-fold, at least or up to about 6-fold, at least or up to about 7-fold, at least or up to about 8-fold, at least or up to about 9-fold, at least or up to about 10-fold, at least or up to about 20-fold, at least or up to about 30-fold, at least or up to about 40-fold, at least or up to about 50-fold, at least or up to about 60-fold, at least or up to about 70-fold, at least or up to about 80-fold, at least or up to about 90-fold, at least or up to about 100-fold, at least or up to about 500-fold, at least or up to about 1,000-fold, at least or up to about 5,000-fold, or at least or up to about 10,000-fold, as compared to a control expression level.

[00250] In some cases, methods disclosed herein can increase expression of a target mRNA by at most or less than about 10,000-fold, at most or less than about 5,000-fold, at most or less than about 1,000-fold, at most or less than about 500-fold, at most or less than about 100-fold, at most or less than about 90-fold, at most or less than about 80-fold, at most or less than about 70-fold, at most or less than about 60-fold, at most or less than about 50-fold, at most or less than about 40-fold, at most or less than about 30-fold, at most or less than about 20-fold, at most or less than about 10-fold, at most or less than about 9-fold, at most or less than about 8-fold, at most or less than about 7-fold, at most or less than about 6-fold, at most or less than about 5-fold, at most or less than about 4-fold, at most or less than about 3-fold, at most or less than about 2-fold, at most or less than about 1-fold, at most or less than about 0.9-fold, at most or less than about 0.8-fold, at most or less than about 0.7-fold, at most or less than about 0.6-fold, at most or less than about 0.5-fold, at most or less than about 0.4-fold, at most or less than about 0.3-fold, at most or less than about 0.2-fold, at most or less than about 0.1-fold, as compared to a control expression level.

[00251] In some cases, methods disclosed herein can decrease expression of a target mRNA by at least about 0.1%, at least about 0.2%, at least about 0.3%, at least about 0.4%, at least about 0.5%, at least about 0.6%, at least about 0.7%, at least about 0.8%, at least about 0.9%, at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 20%, at least about

30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, or more as compared to a control expression level.

[00252] In some cases, methods disclosed herein can decrease expression of a target mRNA by at most about 500%, at most about 400%, at most about 300%, at most about 200%, at most about 100%, at most about 90%, at most about 80%, at most about 70%, at most about 60%, at most about 50%, at most about 40%, at most about 30%, at most about 20%, at most about 10%, at most about 9%, at most about 8%, at most about 7%, at most about 6%, at most about 5%, at most about 4%, at most about 3%, at most about 2%, at most about 1%, at most about 0.9%, at most about 0.8%, at most about 0.7%, at most about 0.6%, at most about 0.5%, at most about 0.4%, at most about 0.3%, at most about 0.2%, at most about 0.1%, or less as compared to a control expression level.

[00253] In some cases, methods disclosed herein can decrease expression of a target mRNA by at least or up to about 0.1-fold, at least or up to about 0.2-fold, at least or up to about 0.3-fold, at least or up to about 0.4-fold, at least or up to about 0.5-fold, at least or up to about 0.6-fold, at least or up to about 0.7-fold, at least or up to about 0.8-fold, at least or up to about 0.9-fold, at least or up to about 1-fold, at least or up to about 2-fold, at least or up to about 3-fold, at least or up to about 4-fold, at least or up to about 5-fold, at least or up to about 6-fold, at least or up to about 7-fold, at least or up to about 8-fold, at least or up to about 9-fold, at least or up to about 10-fold, at least or up to about 20-fold, at least or up to about 30-fold, at least or up to about 40-fold, at least or up to about 50-fold, at least or up to about 60-fold, at least or up to about 70-fold, at least or up to about 80-fold, at least or up to about 90-fold, at least or up to about 100-fold, at least or up to about 500-fold, at least or up to about 1,000-fold, at least or up to about 5,000-fold, or at least or up to about 10,000-fold, as compared to a control expression level.

[00254] In some cases, methods disclosed herein can decrease expression of a target mRNA by at most or less than about 10,000-fold, at most or less than about 5,000-fold, at most or less than about 1,000-fold, at most or less than about 500-fold, at most or less than about 100-fold, at most or less than about 90-fold, at most or less than about 80-fold, at most or less than about 70-fold, at most or less than about 60-fold, at most or less than about 50-fold, at most or less than about 40-fold, at most or less than about 30-fold, at most or less than about 20-fold, at most or less than about 10-fold, at most or less than about 9-fold, at most or less than about 8-fold, at most or less than about 7-fold, at most or less than about 6-fold, at most or less than about 5-fold, at most or less than about 4-fold, at most or less than about 3-fold, at most or less than about 2-fold, at most or less than about 1-fold, at most or less than about 0.9-fold, at most or less than about 0.8-fold, at most or less than about 0.7-fold, at most or less than about 0.6-fold, at most or less than about 0.5-fold, at most or less than about

0.4-fold, at most or less than about 0.3-fold, at most or less than about 0.2-fold, at most or less than about 0.1-fold, as compared to a control expression level.

[00255] A target nucleic acid can be located anywhere, for example, outside of a cell *in vitro*, inside of a cell *in vitro*, inside of a cell *in vivo*, inside of a cell *ex vivo*. Suitable target cells (which can comprise target nucleic acids such as genomic DNA) include, but are not limited to: a bacterial cell; an archaeal cell; a cell of a single-cell eukaryotic organism; a plant cell; an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh*, and the like; a fungal cell (e.g., a yeast cell); an animal cell; a cell from an invertebrate animal (e.g. fruit fly, a cnidarian, an echinoderm, a nematode, etc.); a cell of an insect (e.g., a mosquito; a bee; an agricultural pest; etc.); a cell of an arachnid (e.g., a spider; a tick; etc.); a cell from a vertebrate animal (e.g., a fish, an amphibian, a reptile, a bird, a mammal); a cell from a mammal (e.g., a cell from a rodent; a cell from a human; a cell of a non-human mammal; a cell of a rodent (e.g., a mouse, a rat); a cell of a lagomorph (e.g., a rabbit); a cell of an ungulate (e.g., a cow, a horse, a camel, a llama, a vicuña, a sheep, a goat, etc.); a cell of a marine mammal (e.g., a whale, a seal, an elephant seal, a dolphin, a sea lion; etc.) and the like. Any type of cell may be of interest (e.g. a stem cell, e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell, a germ cell (e.g., an oocyte, a sperm, an oogonia, a spermatogonia, etc.), an adult stem cell, a somatic cell, e.g. a fibroblast, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell; an in vitro or in vivo embryonic cell of an embryo at any stage, e.g., a 1-cell, 2-cell, 4-cell, 8-cell, etc. stage zebrafish embryo; etc.).

[00256] Cells may be from established cell lines or they may be primary cells, where "primary cells", "primary cell lines", and "primary cultures" are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages, i.e. splittings, of the culture. For example, primary cultures are cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times go through the crisis stage. Typically, the primary cell lines are maintained for fewer than 10 passages *in vitro*. Target cells can be unicellular organisms and/or can be grown in culture. If the cells are primary cells, they may be harvest from an individual by any convenient method. For example, leukocytes may be conveniently harvested by apheresis, leukocytapheresis, density gradient separation, etc., while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. can be conveniently harvested by biopsy.

[00257] In some of the above applications, the subject methods may be employed to induce target nucleic acid cleavage, target nucleic acid modification, and/or to bind target nucleic acids (e.g., for visualization, for collecting and/or analyzing, etc.) in mitotic or post-mitotic cells *in vivo* and/or *ex*

vivo and/or *in vitro* (e.g., to disrupt production of a protein encoded by a targeted mRNA, to cleave or otherwise modify target DNA, to genetically modify a target cell, and the like). Because the guide RNA provides specificity by hybridizing to target nucleic acid, a mitotic and/or post-mitotic cell of interest in the disclosed methods may include a cell from any organism (e.g. a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a plant cell, an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa*, *Sargassum patens, C. agardh*, and the like, a fungal cell (e.g., a yeast cell), an animal cell, a cell from an invertebrate animal (e.g. fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal, a cell from a rodent, a cell from a human, etc.). In some cases, a subject CRISPR-Cas effector protein (and/or nucleic acid encoding the protein such as DNA and/or RNA), and/or CRISPR-Cas effector guide RNA (and/or a DNA encoding the guide RNA), and/or donor template, and/or RNP can be introduced into an individual (i.e., the target cell can be in vivo) (e.g., a mammal, a rat, a mouse, a pig, a primate, a non-human primate, a human, etc.). In some case, such an administration can be for the purpose of treating and/or preventing a disease, e.g., by editing the genome of targeted cells.

[00258] Additional examples of target cells can be any cell described herein. Non-limiting examples of cells (target cells) include: a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a protozoa cell, a cell from a plant (e.g., cells from plant crops, fruits, vegetables, grains, soy bean, corn, maize, wheat, seeds, tomatoes, rice, cassava, sugarcane, pumpkin, hay, potatoes, cotton, cannabis, tobacco, flowering plants, conifers, gymnosperms, angiosperms, ferns, clubmosses, hornworts, liverworts, mosses, dicotyledons, monocotyledons, etc.), an algal cell, (e.g., *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh*, and the like), seaweeds (e.g. kelp) a fungal cell (e.g., a yeast cell, a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g., fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal (e.g., an ungulate (e.g., a pig, a cow, a goat, a sheep); a rodent (e.g., a rat, a mouse); a non-human primate; a human; a feline (e.g., a cat); a canine (e.g., a dog); etc.), and the like. In some cases, the cell is a cell that does not originate from a natural organism (e.g., the cell can be a synthetically made cell; also referred to as an artificial cell).

[00259] A cell can be an *in vitro* cell (e.g., established cultured cell line). A cell can be an *ex vivo* cell (cultured cell from an individual). A cell can be and *in vivo* cell (e.g., a cell in an individual). A cell can be an isolated cell. The terms “isolated” and “purified” as used herein can refer to a material that is removed from at least one component with which it is naturally associated (e.g., removed

from its original environment). The term "isolated," when used in reference to an isolated protein, can refer to a protein that has been removed from the culture medium of the host cell that expressed the protein. As such an isolated protein is free of extraneous or unwanted compounds (e.g., nucleic acids, native bacterial or other proteins, etc.).

[00260] A cell can be a cell inside of an organism. A cell can be an organism. A cell can be a cell in a cell culture (e.g., *in vitro* cell culture). A cell can be one of a collection of cells. A cell can be a eukaryotic cell or derived from a eukaryotic cell. A cell can be a plant cell or derived from a plant cell. A cell can be an animal cell or derived from an animal cell. A cell can be an invertebrate cell or derived from an invertebrate cell. A cell can be a vertebrate cell or derived from a vertebrate cell. A cell can be a mammalian cell or derived from a mammalian cell. A cell can be a rodent cell or derived from a rodent cell. A cell can be a human cell or derived from a human cell. A cell can be a fungal cell or derived from a fungal cell. A cell can be an insect cell. A cell can be an arthropod cell. A cell can be a protozoan cell. A cell can be a helminth cell.

[00261] Suitable cells include a stem cell (e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell; a germ cell (e.g., an oocyte, a sperm, an oogonia, a spermatogonia, etc.); a somatic cell, e.g. a fibroblast, an oligodendrocyte, a glial cell, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell, etc.

[00262] Suitable cells include human embryonic stem cells, fetal cardiomyocytes, myofibroblasts, mesenchymal stem cells, cardiomyocytes, adipocytes, totipotent cells, pluripotent cells, blood stem cells, myoblasts, adult stem cells, bone marrow cells, mesenchymal cells, embryonic stem cells, parenchymal cells, epithelial cells, endothelial cells, mesothelial cells, fibroblasts, osteoblasts, chondrocytes, exogenous cells, endogenous cells, stem cells, hematopoietic stem cells, bone-marrow derived progenitor cells, myocardial cells, skeletal cells, fetal cells, undifferentiated cells, multi-potent progenitor cells, unipotent progenitor cells, monocytes, cardiac myoblasts, skeletal myoblasts, macrophages, capillary endothelial cells, xenogeneic cells, allogenic cells, and post-natal stem cells.

[00263] In some cases, the cell is an immune cell, a neuron, an epithelial cell, and endothelial cell, or a stem cell. In some cases, the immune cell is a T cell, a B cell, a monocyte, a natural killer cell, a dendritic cell, or a macrophage. In some cases, the immune cell is a cytotoxic T cell. In some cases, the immune cell is a helper T cell. In some cases, the immune cell is a regulatory T cell (Treg).

[00264] In some cases, the cell is a stem cell. Stem cells include adult stem cells. Adult stem cells are also referred to as somatic stem cells.

[00265] Adult stem cells are resident in differentiated tissue, but retain the properties of self-renewal and ability to give rise to multiple cell types, usually cell types typical of the tissue in which the stem cells are found. Numerous examples of somatic stem cells are known to those of skill in the art,

including muscle stem cells; hematopoietic stem cells; epithelial stem cells; neural stem cells; mesenchymal stem cells; mammary stem cells; intestinal stem cells; mesodermal stem cells; endothelial stem cells; olfactory stem cells; neural crest stem cells; and the like.

[00266] Stem cells of interest include mammalian stem cells, where the term "mammalian" refers to any animal classified as a mammal, including humans; non-human primates; domestic and farm animals; and zoo, laboratory, sports, or pet animals, such as dogs, horses, cats, cows, mice, rats, rabbits, etc. In some cases, the stem cell is a human stem cell. In some cases, the stem cell is a rodent (e.g., a mouse; a rat) stem cell. In some cases, the stem cell is a non-human primate stem cell.

[00267] Stem cells can express one or more stem cell markers, e.g., SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17, and PPARGC1A.

[00268] In some cases, the stem cell is a hematopoietic stem cell (HSC). HSCs are mesoderm-derived cells that can be isolated from bone marrow, blood, cord blood, fetal liver and yolk sac. HSCs are characterized as $CD34^+$ and $CD3^-$. HSCs can repopulate the erythroid, neutrophil-macrophage, megakaryocyte and lymphoid hematopoietic cell lineages *in vivo*. *In vitro*, HSCs can be induced to undergo at least some self-renewing cell divisions and can be induced to differentiate to the same lineages as is seen *in vivo*. As such, HSCs can be induced to differentiate into one or more of erythroid cells, megakaryocytes, neutrophils, macrophages, and lymphoid cells.

[00269] In other cases, the stem cell is a neural stem cell (NSC). Neural stem cells (NSCs) are capable of differentiating into neurons, and glia (including oligodendrocytes, and astrocytes). A neural stem cell is a multipotent stem cell which is capable of multiple divisions, and under specific conditions can produce daughter cells which are neural stem cells, or neural progenitor cells that can be neuroblasts or glioblasts, e.g., cells committed to become one or more types of neurons and glial cells respectively. Methods of obtaining NSCs are known in the art.

[00270] In other cases, the stem cell is a mesenchymal stem cell (MSC). MSCs originally derived from the embryonal mesoderm and isolated from adult bone marrow, can differentiate to form muscle, bone, cartilage, fat, marrow stroma, and tendon. Methods of isolating MSC are known in the art; and any known method can be used to obtain MSC. See, e.g., U.S. Pat. No. 5,736,396, which describes isolation of human MSC.

[00271] A cell is in some cases a plant cell. Plant cells include cells of a monocotyledon, and cells of a dicotyledon. The cells can be root cells, leaf cells, cells of the xylem, cells of the phloem, cells of the cambium, apical meristem cells, parenchyma cells, collenchyma cells, sclerenchyma cells, and the like. Plant cells include cells of agricultural crops such as wheat, corn, rice, sorghum, millet, soybean, etc. Plant cells include cells of agricultural fruit and nut plants, e.g., plant that produce apricots, oranges, lemons, apples, plums, pears, almonds, etc. For example, the cell can be a cell of a

major agricultural plant, e.g., Barley, Beans (Dry Edible), Canola, Corn, Cotton (Pima), Cotton (Upland), Flaxseed, Hay (Alfalfa), Hay (Non-Alfalfa), Oats, Peanuts, Rice, Sorghum, Soybeans, Sugarbeets, Sugarcane, Sunflowers (Oil), Sunflowers (Non-Oil), Sweet Potatoes, Tobacco (Burley), Tobacco (Flue-cured), Tomatoes, Wheat (Durum), Wheat (Spring), Wheat (Winter), and the like. As another example, the cell is a cell of a vegetable crops which include but are not limited to, e.g., alfalfa sprouts, aloe leaves, arrow root, arrowhead, artichokes, asparagus, bamboo shoots, banana flowers, bean sprouts, beans, beet tops, beets, bittermelon, bok choy, broccoli, broccoli rabe (rappini), brussels sprouts, cabbage, cabbage sprouts, cactus leaf (nopales), calabaza, cardoon, carrots, cauliflower, celery, chayote, chinese artichoke (crosnes), chinese cabbage, chinese celery, chinese chives, choy sum, chrysanthemum leaves (tung ho), collard greens, corn stalks, corn-sweet, cucumbers, daikon, dandelion greens, dasheen, dau mue (pea tips), donqua (winter melon), eggplant, endive, escarole, fiddle head ferns, field cress, frisee, gai choy (chinese mustard), gailon, galanga (siam, thai ginger), garlic, ginger root, gobo, greens, hanover salad greens, huauzontle, jerusalem artichokes, jicama, kale greens, kohlrabi, lamb's quarters (quilete), lettuce (bibb), lettuce (boston), lettuce (boston red), lettuce (green leaf), lettuce (iceberg), lettuce (lolla rossa), lettuce (oak leaf - green), lettuce (oak leaf - red), lettuce (processed), lettuce (red leaf), lettuce (romaine), lettuce (ruby romaine), lettuce (russian red mustard), linkok, lo bok, long beans, lotus root, mache, maguey (agave) leaves, malanga, mesculin mix, mizuna, moap (smooth luffa), moo, moqua (fuzzy squash), mushrooms, mustard, nagaimo, okra, ong choy, onions green, opo (long squash), ornamental corn, ornamental gourds, parsley, parsnips, peas, peppers (bell type), peppers, pumpkins, radicchio, radish sprouts, radishes, rape greens, rape greens, rhubarb, romaine (baby red), rutabagas, salicornia (sea bean), sinqua (angled/ridged luffa), spinach, squash, straw bales, sugarcane, sweet potatoes, swiss chard, tamarindo, taro, taro leaf, taro shoots, tatsoi, tepeguaje (guaje), tindora, tomatillos, tomatoes, tomatoes (cherry), tomatoes (grape type), tomatoes (plum type), tumeric, turnip tops greens, turnips, water chestnuts, yampi, yams, yu choy, yuca (cassava), and the like.

[00272] A cell is in some cases an arthropod cell. For example, the cell can be a cell of a sub-order, a family, a sub-family, a group, a sub-group, or a species of, e.g., *Chelicerata*, *Myriapodia*, *Hexipodia*, *Arachnida*, *Insecta*, *Archaeognatha*, *Thysanura*, *Palaeoptera*, *Ephemeroptera*, *Odonata*, *Anisoptera*, *Zygoptera*, *Neoptera*, *Exopterygota*, *Plecoptera*, *Embioptera*, *Orthoptera*, *Zoraptera*, *Dermaptera*, *Dictyoptera*, *Notoptera*, *Grylloblattidae*, *Mantophasmatidae*, *Phasmatodea*, *Blattaria*, *Isoptera*, *Mantodea*, *Parapneuroptera*, *Psocoptera*, *Thysanoptera*, *Phthiraptera*, *Hemiptera*, *Endopterygota or Holometabola*, *Hymenoptera*, *Coleoptera*, *Strepsiptera*, *Raphidioptera*, *Megaloptera*, *Neuroptera*, *Mecoptera*, *Siphonaptera*, *Diptera*, *Trichoptera*, or *Lepidoptera*.

[00273] A cell is in some cases an insect cell. For example, in some cases, the cell is a cell of a mosquito, a grasshopper, a true bug, a fly, a flea, a bee, a wasp, an ant, a louse, a moth, or a beetle.

KIT

[00274] The present disclosure provides kits comprising one or more of the compositions disclosed herein (e.g., polynucleotide, polypeptide, etc.), one or more of the nucleic acids disclosed herein, one or more of the expression vectors disclosed herein, or the cells disclosed herein; and instructions for use. In some embodiments, the instructions for use can direct operation of methods of editing a locus within a cell or modifying a target nucleic acid described herein.

PHARMACEUTICAL COMPOSITIONS

[00275] The present disclosure provides pharmaceutical compositions comprising one or more of the polynucleotides disclosed herein (e.g., fusion polynucleotide), one or more of the polypeptides disclosed herein (e.g., fusion polypeptide), one or more of the expression vectors disclosed herein, or the cells disclosed herein, and a pharmaceutically acceptable carrier, excipient, or diluent. In some embodiments, the compositions of the disclosure can be incorporated into a pharmaceutical composition. Such compositions typically include the compositions of the disclosure (e.g., a fusion polynucleotide, a fusion polypeptide, etc.), and a pharmaceutically acceptable excipient, e.g., a carrier. Compositions of the present disclosure can be administered using formulations used for administering antibodies and antibody-based therapeutics, or formulations based thereon.

[00276] Pharmaceutical compositions or formulations comprising the polynucleotides, polypeptides, or any combination thereof (e.g., fusion polynucleotide, fusion polypeptide, etc.), of the described compositions and for use in any of the described methods can be prepared according to conventional techniques well known in the pharmaceutical industry and described in the published literature. In some embodiments, a pharmaceutical composition or formulation for treating a subject comprises an effective amount of any polynucleotides, polypeptides, or combination thereof (e.g., fusion polynucleotide, fusion polypeptide, etc.), of the described compositions, or a pharmaceutically acceptable salt, solvate, hydrate or ester thereof. The pharmaceutical formulation comprising the polynucleotides, polypeptides, or any combination thereof (e.g., fusion polynucleotide, fusion polypeptide, etc.), of the described compositions may further comprise a pharmaceutically acceptable excipient, diluent or carrier.

[00277] Pharmaceutically acceptable salts are suitable for use in contact with the tissues of humans and lower animals without undue toxicity, irritation, allergic response, etc., and are commensurate with a reasonable benefit/risk ratio. (See, e.g., S. M. Berge, et al., J. Pharmaceutical Sciences, 66: 1-

19 (1977), incorporated herein by reference for this purpose. The salts can be prepared in situ during the final isolation and purification of the compounds, or separately by reacting the free base form with a suitable organic acid. Examples of pharmaceutically acceptable, nontoxic acid addition salts are salts of an amino group formed with inorganic acids such as hydrochloric acid, hydrobromic acid, phosphoric acid, sulfuric acid and perchloric acid or with organic acids such as acetic acid, oxalic acid, maleic acid, tartaric acid, citric acid, succinic acid or malonic acid or by using other documented methodologies such as ion exchange. Other pharmaceutically acceptable salts include adipate, alginate, ascorbate, aspartate, benzenesulfonate, benzoate, bisulfate, borate, butyrate, camphorate, camphorsulfonate, citrate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, formate, fumarate, glucoheptonate, glycerophosphate, gluconate, hemisulfate, heptanoate, hexanoate, hydroiodide, 2-hydroxy-ethanesulfonate, lactobionate, lactate, laurate, lauryl sulfate, malate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oleate, oxalate, palmitate, pamoate, pectinate, persulfate, 3-phenylpropionate, phosphate, picrate, pivalate, propionate, stearate, succinate, sulfate, tartrate, thiocyanate, p-toluenesulfonate, undecanoate, valerate salts, and the like. Representative alkali or alkaline earth metal salts include sodium, lithium, potassium, calcium, magnesium, and the like. Further pharmaceutically acceptable salts include, when appropriate, nontoxic ammonium, quaternary ammonium, and amine cations formed using counterions such as halide, hydroxide, carboxylate, sulfate, phosphate, nitrate, loweralkyl sulfonate and aryl sulfonate.

[00278] In some embodiments, the compositions are formulated as suspensions in aqueous, non-aqueous or mixed media. Aqueous suspensions may further contain substances that increase the viscosity of the suspension including, for example, sodium carboxymethylcellulose, sorbitol and/or dextran. The suspension may also contain stabilizers. In some embodiments, a pharmaceutical formulation or composition as described herein includes, but is not limited to, a solution, emulsion, microemulsion, foam or liposome-containing formulation (e.g., cationic or noncationic liposomes).

[00279] The pharmaceutical composition or formulation described herein may comprise one or more penetration enhancers, carriers, excipients or other active or inactive ingredients as appropriate and well known to those of skill in the art or described in the published literature. In some embodiments, liposomes also include sterically stabilized liposomes, e.g., liposomes comprising one or more specialized lipids. These specialized lipids result in liposomes with enhanced circulation lifetimes. In some embodiments, a sterically stabilized liposome comprises one or more glycolipids or is derivatized with one or more hydrophilic polymers, such as a polyethylene glycol (PEG) moiety. In some embodiments, a surfactant is included in the pharmaceutical formulation or compositions. The use of surfactants in drug products, formulations and emulsions is well known in the art. In some

embodiments, the present disclosure employs a penetration enhancer to effect the efficient delivery of the polynucleotides, polypeptides, or any combination thereof (e.g., fusion polynucleotide, fusion polypeptide, etc.), of the described compositions, e.g., to aid diffusion across cell membranes and /or enhance the permeability of a lipophilic drug. In some embodiments, the penetration enhancers are a surfactant, fatty acid, bile salt, chelating agent, or non-chelating nonsurfactant.

METHODS OF TREATMENT

[00280] The present disclosure provides methods of treating a subject with a condition or a disease. In some embodiments, the method comprises administering to the subject a therapeutically effective amount of the pharmaceutical composition disclosed herein. In some cases, a method of modifying a target nucleic acid provides for treatment of a disease or disorder in a subject. The present disclosure also provides a method of treating a disease or disorder in a subject, the method comprising administering to the individual a fusion polynucleotide of the present disclosure. The present disclosure also provides a method of treating a disease or disorder in a subject, the method comprising administering to the individual: a) a fusion polypeptide of the present disclosure; or b) a fusion polypeptide of the present disclosure and a guide nucleic acid; or c) a fusion polypeptide of the present disclosure, a guide nucleic acid, and a donor nucleic acid. The terms "individual," "subject," "host," and "patient," can be used interchangeably herein, and may refer to an individual organism, e.g., a mammal, including, but not limited to, murines, simians, humans, non-human primates, ungulates, felines, canines, bovines, ovines, mammalian farm animals, mammalian sport animals, and mammalian pets.

[00281] As used herein, the terms "treatment," "treating," and the like, may refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. "Treatment," as used herein, covers any treatment of a disease in a mammal, e.g., in a human, and includes: (a) preventing the disease from occurring in a subject which may be predisposed to the disease but has not yet been diagnosed as having it; (b) inhibiting the disease, i.e., arresting its development; and (c) relieving the disease, i.e., causing regression of the disease.

[00282] In some embodiments, the condition or disease is a myopathy, a nephropathy, or a brain disease.

[00283] In some embodiments, the condition or disease is a myopathy. In some embodiments, the condition or disease is selected from the group consisting of: Duchenne Muscular Dystrophy, Myotonic Dystrophy Type 1, Myotonic Dystrophy Type 2, Pompe Disease, Cardiomyopathy (e.g., Hypertrophic cardiomyopathy (MYBPC3), Desmoplakin cardiomyopathy, Dilated cardiomyopathy,

Arrhythmogenic Cardiomyopathy, Plakophilin-2 Related Arrhythmogenic Cardiomyopathy, etc.), Hutchinson–Gilford Progeria Syndrome, Congenital Myasthenic Syndromes (e.g., Collagen Q Congenital Myasthenic Syndrome), Heart Failure (e.g., Heart failure with reduced ejection fraction, Heart failure with cardiac hypertrophy, Heart failure due to myocardial ischemia, Congestive heart failure, Systolic heart failure, etc.), Sarcopenia (e.g., Aging-Related Sarcopenia), Muscular Atrophy (e.g., Skeletal muscle atrophy), Facioscapulohumeral muscular dystrophy, Friedreich's Ataxia, Limb-Girdle Muscular Dystrophy (e.g., LGMD2B, LGMD2D, LGMD2E, LAMA2-related muscular dystrophy, etc.), Glycogen Storage Disorders, Danon Disease, X-linked Mytotubular Myopathy, TTR Amyloidosis, Alpha-1 Antitrypsin Deficiency, Tyrosinemia, Phenylketonuria, Propionic Acidemia, Primary Hyperoxaluria, Acute Intermittent Porphyria, Hypertriglyceridemia, and Methylmalonic Acidemia. In some embodiments, the targeting moiety (as disclosed herein) binds to a TfR1 receptor.

[00284] In some embodiments, the condition or disease is a brain disease. In some embodiments, the condition or disease is selected from the group consisting of: Dravet syndrome, Angelman Syndrome, Developmental and Epileptic Encephalopathy, Epilepsies (e.g., Temporal Lobe Epilepsy, STXBP1-mediated epilepsy, etc.), Amyotrophic lateral sclerosis, Huntington's disease, Parkinson's disease, Alzheimer's Disease, Spinal Muscular Atrophy, Frontotemporal Dementia, Spinocerebellar Ataxia, Schizophrenia, Borderline personality disorder, Kabuki syndrome, Autism spectrum disorder, Fragile X Syndrome, Adrenomyeloneuropathy, Progressive Supranuclear Palsy, Canavan Disease, Seizures (e.g., CDKL5 deficiency disorder), CLN1 Disease, Batten Disease, Aromatic l-Amino Acid Decarboxylase Deficiency, Friedrich's ataxia, Krabbe Disease, Multiple System Atrophy, GM1 Gangliosidosis, Hunter Syndrome, Hurler Syndrome, Sanfilippo Syndrome, Mucopolysaccharidosis Type IIIB, Rett Syndrome, a- Synucleinopathies, Alexander disease, Leukodystrophies, and pain (e.g., Neuropathic pain, Chronic pain, Acute Pain, etc.). In some embodiments, the targeting moiety (as disclosed herein) binds to an IGF1R receptor.

[00285] In some embodiments, the condition or disease is a nephropathy, or a brain disease. In some embodiments, the condition or disease is selected from the group consisting of: Dravet syndrome, Angelman Syndrome, Developmental and Epileptic Encephalopathy, Epilepsies (e.g., Temporal Lobe Epilepsy, STXBP1-mediated epilepsy, etc.), Amyotrophic lateral sclerosis, Huntington's disease, Parkinson's disease, Alzheimer's Disease, Spinal Muscular Atrophy, Frontotemporal Dementia, Spinocerebellar Ataxia, Schizophrenia, Borderline personality disorder, Kabuki syndrome, Autism spectrum disorder, Fragile X Syndrome, Adrenomyeloneuropathy, Progressive Supranuclear Palsy, Canavan Disease, Seizures (e.g., CDKL5 deficiency disorder), CLN1 Disease, Batten Disease, Aromatic l-Amino Acid Decarboxylase Deficiency, Friedrich's ataxia, Krabbe

Disease, Multiple System Atrophy, GM1 Gangliosidosis, Hunter Syndrome, Hurler Syndrome, Sanfilippo Syndrome, Mucopolysaccharidosis Type IIIB, Rett Syndrome, a- Synucleinopathies, Alexander disease, Leukodystrophies, Pain (e.g., Neuropathic pain, Chronic pain, Acute Pain, etc.), Chronic kidney disease, APOL1-mediated kidney disease, Autosomal Dominant Polycystic Kidney Disease, Autosomal Recessive Polycystic Kidney Disease, Alport syndrome, Kidney Fibrosis (e.g., Tubulo-interstitial fibrosis), Liddle syndrome, Corticosteroid remediable aldosteronism, Familial hyperaldosteronism (type II), Familial hyperaldosteronism (type III), Fabry disease, COQ10 deficiency, Amyloidosis, Focal segmental glomerulosclerosis, Cystinuria, Cystinosis, Tuberous Sclerosis, Wolfram syndrome, Distal renal tubular acidosis, Autosomal Dominant Polycystic Kidney Disease, Glomerulonephritis, Bartter Syndrome, Gitelman Syndrome, Acute Kidney Injury (e.g., Delayed graft function after surgery, Ischemic and cisplatin-induced kidney injury, Renal ischemia reperfusion injury, Acute kidney injury after unilateral urethral obstruction, etc.), Immunoglobulin A nephropathy, Phenylketonuria, Primary hyperoxaluria, and Pseudohyperaldosteronism (type II). In some embodiments, the targeting moiety (as disclosed herein) binds to an IGF2R receptor.

[00286] In some embodiments, the condition or disease is a nephropathy. In some embodiments, the condition or disease is selected from the group consisting of: Chronic kidney disease, APOL1-mediated kidney disease, Autosomal Dominant Polycystic Kidney Disease, Autosomal Recessive Polycystic Kidney Disease, Alport syndrome, Kidney Fibrosis (e.g., Tubulo-interstitial fibrosis), Liddle syndrome, Corticosteroid remediable aldosteronism, Familial hyperaldosteronism (type II), Familial hyperaldosteronism (type III), Fabry disease, COQ10 deficiency, Amyloidosis, Focal segmental glomerulosclerosis, Cystinuria, Cystinosis, Tuberous Sclerosis, Wolfram syndrome, Distal renal tubular acidosis, Autosomal Dominant Polycystic Kidney Disease, Glomerulonephritis, Bartter Syndrome, Gitelman Syndrome, Acute Kidney Injury (e.g., Delayed graft function after surgery, Ischemic and cisplatin-induced kidney injury, Renal ischemia reperfusion injury, Acute kidney injury after unilateral urethral obstruction, etc.), Immunoglobulin A nephropathy, Phenylketonuria, Primary hyperoxaluria, and Pseudohyperaldosteronism (type II). In some embodiments, the targeting moiety (as disclosed herein) binds to a Nephrin receptor.

[00287] Any of the compositions provided herein may be administered to an individual. "Individual" may be used interchangeably with "subject" or "patient." An individual may be a mammal, for example a human or animal such as a non-human primate, a rodent, a rabbit, a rat, a mouse, a horse, a donkey, a goat, a cat, a dog, a cow, a pig, or a sheep. In some embodiments, the individual is a human. In some embodiments, the individual is a fetus, an embryo, or a child. In other embodiments, the individual may be another eukaryotic organism, such as a plant. In some embodiments, the compositions provided herein are administered to a cell ex vivo.

[00288] In some embodiments, the compositions provided herein are administered to an individual as a method of treating a disease or disorder. In some embodiments, the individual has a genetic disease, such as any of the diseases described herein. In some embodiments, the individual is at risk of having a disease, such as any of the diseases described herein. In some embodiments, the individual is at increased risk of having a disease or disorder caused by insufficient amount of a protein or insufficient activity of a protein. In some embodiments, the individual is at increased risk of having a disease or disorder caused by an excessive amount of a protein or excessive activity of a protein. In some embodiments, the individual is at increased risk of having a disease or disorder caused by abnormal activity of a protein. If an individual is "at an increased risk" of having a disease or disorder, the method involves preventative or prophylactic treatment. For example, an individual may be at an increased risk of having such a disease or disorder because of family history of the disease. Typically, individuals at an increased risk of having such a disease or disorder benefit from prophylactic treatment (e.g., by preventing or delaying the onset or progression of the disease or disorder). In some embodiments, a fetus is treated in utero, e.g., by administering the pharmaceutical compositions or the compositions as described herein to the fetus directly or indirectly (e.g., via the mother).

[00289] Suitable routes for administration of the pharmaceutical compositions or the compositions as described herein may vary depending on cell type to which delivery of the pharmaceutical compositions or the compositions is desired. The pharmaceutical compositions or the compositions as described herein may be administered to patients parenterally, for example, by intrathecal injection, intracerebroventricular injection, intraperitoneal injection, intramuscular injection, subcutaneous injection, or intravenous injection.

DEFINITIONS

[00290] The following definitions supplement those in the art and are directed to the current application and are not to be imputed to any related or unrelated case, e.g., to any commonly owned patent or application. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the present disclosure, the preferred materials and methods are described herein. Accordingly, the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

[00291] The use of the terms "a," "an," and "the," and similar referents in the context of describing the disclosure (especially in the context of the following claims) may be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms "comprising," "having," "including," and "containing" are to be construed as open-ended terms (*i.e.*, meaning "including, but not limited to,") unless otherwise noted. Recitation of ranges of

values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. For example, if the range 10-15 is disclosed, then 11, 12, 13, and 14 are also disclosed. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (*e.g.*, "such as") provided herein, is intended merely to better illuminate the embodiments of the disclosure and does not pose a limitation on the scope of the disclosure unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the embodiments of the disclosure.

[00292] In this application, the use of "and/or" as used herein a phrase such as "A and/or B" is intended to include both A and B; A or B; A (alone); and B (alone). Likewise, the term "and/or" as used herein a phrase such as "A, B, and/or C" is intended to encompass each of the following embodiments: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

[00293] Reference in the specification to "some embodiments," "an embodiment," "one embodiment" or "other embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments, of the inventions.

[00294] As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method or composition of the invention, and vice versa. Furthermore, compositions of the invention can be used to achieve methods of the invention.

[00295] As used herein, the term "about" used in connection with an amount indicates that the amount can vary by 10% of the stated amount. For example, "about 100" means an amount of from 90-110. Where about is used in the context of a range, the "about" used in reference to the lower amount of the range means that the lower amount includes an amount that is 10% lower than the lower amount of the range, and "about" used in reference to the higher amount of the range means that the higher amount includes an amount 10% higher than the higher amount of the range. For example, "from about 100 to about 1000" means that the range extends from 90 to 1100.

[00296] The term "sequence identity" or "percent identity" in the context of two or more nucleic acids or polypeptide sequences, generally refers to two (e.g., in a pairwise alignment) or more (e.g., in a multiple sequence alignment) sequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a local or global comparison window, as measured using a sequence comparison algorithm. Suitable sequence comparison algorithms for polypeptide sequences include, e.g., BLASTP using parameters of a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment for polypeptide sequences longer than 30 residues; BLASTP using parameters of a wordlength (W) of 2, an expectation (E) of 1000000, and the PAM30 scoring matrix setting gap costs at 9 to open gaps and 1 to extend gaps for sequences of less than 30 residues (these are the default parameters for BLASTP in the BLAST suite available at https://blast.ncbi.nlm.nih.gov); CLUSTALW with parameters of; the Smith-Waterman homology search algorithm with parameters of a match of 2, a mismatch of -1, and a gap of -1; MUSCLE with default parameters; MAFFT with parameters retree of 2 and maxiterations of 1000; Novafold with default parameters; HMMER hmmalign with default parameters. Another alignment algorithm is FASTA, available in the Genetics Computing Group (GCG) package, from Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc. Other techniques for alignment are described in Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA. Of particular interest are alignment programs that permit gaps in the sequence. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Meth. Mol. Biol.* 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. See *J. Mol. Biol.* 48: 443-453 (1970).

[00297] Ranges: throughout this disclosure, various aspects of the invention can be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 2.7, 3, 4, 5, 5.3, and 6. As another example, a range such as 95-99% identity, includes something with 95%, 96%, 97%, 98% or 99% identity, and includes subranges such as 96-99%, 96-98%, 96-97%.

[00298] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[00299] All terms are intended to be understood as they would be understood by a person skilled in the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the disclosure pertains.

[00300] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the claimed subject matter belongs. It is to be understood that the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of any subject matter claimed. In this application, the use of the singular includes the plural unless specifically stated otherwise.

NUMBERED EMBODIMENTS

Numbered embodiment 1. A composition comprising a fusion polypeptide comprising:
(a) an effector polypeptide; and
(b) a targeting moiety that specifically binds to an extracellular domain of a receptor expressed by a target cell, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody.

Numbered embodiment 2. The composition of numbered embodiment 1, wherein the fusion polypeptide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

Numbered embodiment 3. The composition of numbered embodiment 1 or 2, wherein the fusion polypeptide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

Numbered embodiment 4. A composition comprising a fusion polypeptide comprising:
(a) an effector polypeptide; and
(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 31-151 or 350-143854.

Numbered embodiment 5. The composition of numbered embodiment 4, wherein the fusion polypeptide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein

when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

Numbered embodiment 6. The composition of numbered embodiment 4 or 5, wherein the fusion polypeptide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

Numbered embodiment 7. A composition comprising a fusion polypeptide comprising:

(a) an effector polypeptide; and

(b) a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

Numbered embodiment 8. A composition comprising a fusion polypeptide comprising:

(a) an effector polypeptide; and

(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 336-349.

Numbered embodiment 9. The composition of numbered embodiment 7 or 8, wherein the fusion polypeptide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell.

Numbered embodiment 10. The composition of numbered embodiment 9, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody.

Numbered embodiment 11. A composition comprising a fusion polypeptide comprising:

(a) an effector polypeptide; and

(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 152-335.

Numbered embodiment 12. A composition comprising a protein complex, the protein complex comprising:

(a) the fusion polypeptide of the composition of numbered embodiment 7 or 8, and

(b) an endosomal escape peptide, wherein the endosomal escape peptide is bound to the fusion polypeptide, and wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

Numbered embodiment 13. The composition of numbered embodiment 11 or 12, wherein the fusion polypeptide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell.

Numbered embodiment 14. The composition of numbered embodiment 13, wherein the endosomal escape peptide is bound to a pH responsive polypeptide domain.

Numbered embodiment 15. The composition of any one of numbered embodiments 1-3, 9-10 and 13-14, wherein the targeting moiety binds to an extracellular domain of transferrin receptor protein 1 (TfR1), IGF1R, IGF2R, Dnmt3a, or Nephrin.

Numbered embodiment 16. The composition of any one of numbered embodiments 1-3, 9-10 and 13-15, wherein the targeting moiety comprises a polypeptide sequence with less than 150 amino acids.

Numbered embodiment 17. The composition of any one of numbered embodiments 1-3, 9-10 and 13-16, wherein the targeting moiety comprises a polypeptide sequence with less than 100 amino acids.

Numbered embodiment 18. The composition of numbered embodiment 16 or 17, wherein the targeting moiety comprises a polypeptide sequence with more than 50 amino acids.

Numbered embodiment 19. The composition of numbered embodiment 18, wherein the targeting moiety comprises a polypeptide sequence with more than 60 amino acids.

Numbered embodiment 20. The composition of numbered embodiment 18, wherein the targeting moiety comprises a polypeptide sequence with more than 70 amino acids.

Numbered embodiment 21. The composition of numbered embodiment 18, wherein the targeting moiety comprises a polypeptide sequence with 80 amino acids or more.

Numbered embodiment 22. The composition of any one of numbered embodiments 1-3, 9-10 and 13-21, wherein the targeting moiety comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence any one of SEQ ID NOs: 31-151 or 350-143854.

Numbered embodiment 23. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated at a pH of 6.8 or less, and wherein the one or more amino acids that are protonated at a pH of 6.8 or less are not protonated at a pH of 7.4.

Numbered embodiment 24. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 6.6 or less, and wherein the one or more amino acids that are protonated at a pH of 6.6 or less are not protonated at a pH of 7.4.

Numbered embodiment 25. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated

at a pH of 6 or less, and wherein the one or more amino acids that are protonated at a pH of 6 or less are not protonated at a pH of 7.4.

Numbered embodiment 26. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5.5 or less, and wherein the one or more amino acids that are protonated at a pH of 5.5 or less are not protonated at a pH of 7.4.

Numbered embodiment 27. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5 or less, and wherein the one or more amino acids that are protonated at a pH of 5 or less are not protonated at a pH of 7.4.

Numbered embodiment 28. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-27, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polypeptide is present within an endosome.

Numbered embodiment 29. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-28, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polypeptide is present within a lysosome.

Numbered embodiment 30. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-29, wherein the pH responsive polypeptide domain comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 336-349.

Numbered embodiment 31. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 32. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 33. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 34. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive

polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 35. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 36. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 37. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 38. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 39. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 40. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-30, wherein at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 41. The composition of any one of numbered embodiments 3, 6, and 12-40, wherein the endosomal escape peptide is released from the protein complex when the protein complex is present within an endosome.

Numbered embodiment 42. The composition of any one of numbered embodiments 3, 6, and 12-41, wherein the endosomal escape peptide is released from the protein complex when the protein complex is present within a lysosome.

Numbered embodiment 43. The composition of any one of numbered embodiments 3, 6, and 12-42, wherein the endosomal escape peptide comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 152-335.

Numbered embodiment 44. The composition of any one of numbered embodiments 2, 3, 5-7, and 14-43, wherein the pH responsive polypeptide domain is capable of binding to two or more endosomal escape peptides at pH 7.4.

Numbered embodiment 45. The composition of any one of numbered embodiments 12-44, wherein the protein complex comprises two or more copies of the endosomal escape peptide.

Numbered embodiment 46. The composition of any one of numbered embodiments 1-45, wherein the fusion polypeptide further comprises a linker linking the effector polypeptide and one or more other peptide sequences of the fusion polypeptide.

Numbered embodiment 47. The composition of numbered embodiment 46, wherein the linker is a proteolytically cleavable linker.

Numbered embodiment 48. The composition of any one of numbered embodiments 1-47, wherein the fusion polypeptide further comprises one or more nuclear localization sequences (NLSs).

Numbered embodiment 49. The composition of numbered embodiment 48, wherein the one or more NLSs comprise the amino acid sequence K(K/R)X(K/R), and where X is any amino acid.

Numbered embodiment 50. The composition of numbered embodiment 48, wherein the one or more NLSs comprise the amino acid sequence PKKKRKV.

Numbered embodiment 51. The composition of any one of numbered embodiments 1-50, wherein the one or more NLSs are at the N-terminus of the effector polypeptide.

Numbered embodiment 52. The composition of any one of numbered embodiments 1-51, wherein the effector polypeptide is a CRISPR-Cas effector polypeptide.

Numbered embodiment 53. The composition of any one of numbered embodiments 1-51, wherein the CRISPR-Cas effector polypeptide is a Type II CRISPR-Cas effector polypeptide, a Type V CRISPR-Cas effector polypeptide, or a Type VI CRISPR-Cas effector polypeptide.

Numbered embodiment 54. The composition of any one of numbered embodiments 1-51, wherein the CRISPR-Cas effector polypeptide is a SpyCas9 or a variant thereof.

Numbered embodiment 55. The composition of any one of numbered embodiments 1-51, wherein the CRISPR-Cas effector polypeptide is GeoCas9 or a variant thereof.

Numbered embodiment 56. The composition of any one of numbered embodiments 1-55, wherein the CRISPR-Cas effector polypeptide is catalytically active.

Numbered embodiment 57. The composition of any one of numbered embodiments 1-56, wherein the CRISPR-Cas effector polypeptide exhibits reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide.

Numbered embodiment 58. The composition of any one of numbered embodiments 1-55, wherein the CRISPR-Cas effector polypeptide is catalytically inactive.

Numbered embodiment 59. The composition of any one of numbered embodiments 1-58, wherein the fusion polypeptide further comprises at least one additional heterologous polypeptide.

Numbered embodiment 60. The composition of numbered embodiment 59, wherein the at least one additional heterologous polypeptide is a deaminase, a base editor, a reverse transcriptase, a transcription modulator, or an epigenetic modulator.

Numbered embodiment 61. A composition comprising:

(a) a fusion polypeptide of the composition of any one of numbered embodiments 1-60, or a nucleic acid comprising a nucleotide sequence encoding the fusion polypeptide and/or endosomal escape polypeptide; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

Numbered embodiment 62. The composition of numbered embodiment 61, wherein the guide nucleic acid is a single-molecule guide nucleic acid.

Numbered embodiment 63. The composition of numbered embodiment 61, further comprising a donor nucleic acid.

Numbered embodiment 64. A nucleic acid comprising a nucleotide sequence encoding the fusion polypeptide of the composition of any one of numbered embodiments 1-60.

Numbered embodiment 65. The nucleic acid of numbered embodiment 64, wherein the nucleotide sequence is operably linked to a transcriptional control element, optionally wherein the transcriptional control element is a promoter.

Numbered embodiment 66. A recombinant expression vector comprising the nucleic acid of numbered embodiment 65.

Numbered embodiment 67. A cell comprising the composition of any one of numbered embodiments 1-63 or the nucleic acid of numbered embodiment 64 or 65, or the recombinant expression vector of numbered embodiment 66.

Numbered embodiment 68. The cell of numbered embodiment 67, wherein the cell is a eukaryotic cell.

Numbered embodiment 69. The cell of numbered embodiment 67 or 68, wherein the cell is in vitro.

Numbered embodiment 70. The cell of numbered embodiment 67 or 68, wherein the cell is in vivo.

Numbered embodiment 71. A method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with:

(a) the composition of any one of numbered embodiments 1-63; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

Numbered embodiment 72. A method of modifying a target nucleic acid in a eukaryotic cell, the method comprising introducing into the eukaryotic cell:

(a) the composition of any one of numbered embodiments 1-63; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

Numbered embodiment 73. The method of numbered embodiment 72, further comprising introducing into the cell a donor nucleic acid.

Numbered embodiment 74. The method of numbered embodiment 72 or 73, wherein the cell is in vitro.

Numbered embodiment 75. The method of numbered embodiment 72 or 73, wherein the cell is in vivo.

Numbered embodiment 76. The method of any one of numbered embodiments 72-75, wherein the cell is a mammalian cell, an insect cell, an avian cell, a reptile cell, an amphibian cell, an arachnid cell, a protozoan cell, or a plant cell.

Numbered embodiment 77. The method of any one of numbered embodiments 72-76, wherein the target nucleic acid is selected from: double stranded DNA, single stranded DNA, RNA, genomic DNA, and extrachromosomal DNA.

Numbered embodiment 78. The method of any one of numbered embodiments 72-77, wherein the modifying comprises genome editing.

Numbered embodiment 79. A composition comprising a fusion polynucleotide comprising:

(a) a polynucleotide; and

(b) a targeting moiety that specifically binds to an extracellular domain of a receptor expressed by a target cell, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody.

Numbered embodiment 80. The composition of numbered embodiment 79, wherein the fusion polynucleotide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein

when the fusion polynucleotide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

Numbered embodiment 81. The composition of numbered embodiment 79 or 80, wherein the fusion polynucleotide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

Numbered embodiment 82. A composition comprising a fusion polynucleotide comprising:

(a) a polynucleotide; and

(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 31-151 or 350-143854.

Numbered embodiment 83. The composition of numbered embodiment 82, wherein the fusion polynucleotide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polynucleotide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

Numbered embodiment 84. The composition of numbered embodiment 82 or 83, wherein the fusion polynucleotide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.

Numbered embodiment 85. A composition comprising a fusion polynucleotide comprising:

(a) a polynucleotide; and

(b) a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polynucleotide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

Numbered embodiment 86. A composition comprising a fusion polynucleotide comprising:

(a) a polynucleotide; and

(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 336-349.

Numbered embodiment 87. The composition of numbered embodiment 85 or 86, wherein the fusion polynucleotide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell.

Numbered embodiment 88. The composition of numbered embodiment 87, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody.

Numbered embodiment 89. A composition comprising a fusion polynucleotide comprising:
(a) a polynucleotide; and
(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 152-335.
Numbered embodiment 90. A composition comprising a protein complex, the protein complex comprising:
(a) the fusion polynucleotide of the composition of numbered embodiment 88 or 89, and
(b) an endosomal escape peptide, wherein the endosomal escape peptide is bound to the fusion polynucleotide, and wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.
Numbered embodiment 91. The composition of numbered embodiment 89 or 90, wherein the fusion polynucleotide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell.
Numbered embodiment 92. The composition of numbered embodiment 91, wherein the endosomal escape peptide is bound to a pH responsive polypeptide domain.
Numbered embodiment 93. The composition of any one of numbered embodiments 79-81, 87, 88, 91, and 92, wherein the targeting moiety binds to an extracellular domain of transferrin receptor protein 1 (TfR1), IGF1R, IGF2R, Dnmt3a, or Nephrin.
Numbered embodiment 94. The composition of any one of numbered embodiments 79-81, 87, 88, and 91-93, wherein the targeting moiety comprises a polypeptide sequence with less than 150 amino acids.
Numbered embodiment 95. The composition of any one of numbered embodiments 79-81, 87, 88, and 91-94, wherein the targeting moiety comprises a polypeptide sequence with less than 100 amino acids.
Numbered embodiment 96. The composition of numbered embodiment 94 or 95, wherein the targeting moiety comprises a polypeptide sequence with more than 50 amino acids.
Numbered embodiment 97. The composition of numbered embodiment 96, wherein the targeting moiety comprises a polypeptide sequence with more than 60 amino acids.
Numbered embodiment 98. The composition of numbered embodiment 96, wherein the targeting moiety comprises a polypeptide sequence with more than 70 amino acids.
Numbered embodiment 99. The composition of numbered embodiment 96, wherein the targeting moiety comprises a polypeptide sequence with 80 amino acids or more.

Numbered embodiment 100. The composition of any one of numbered embodiments 79-81, 87, 88 and 91-99, wherein the targeting moiety comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence any one of SEQ ID NOs: 31-151 or 350-143854.

Numbered embodiment 101. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-100, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated at a pH of 6.8 or less, and wherein the one or more amino acids that are protonated at a pH of 6.8 or less are not protonated at a pH of 7.4.

Numbered embodiment 102. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-100, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 6.6 or less, and wherein the one or more amino acids that are protonated at a pH of 6.6 or less are not protonated at a pH of 7.4.

Numbered embodiment 103. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-100, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 6 or less, and wherein the one or more amino acids that are protonated at a pH of 6 or less are not protonated at a pH of 7.4.

Numbered embodiment 104. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-100, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5.5 or less, and wherein the one or more amino acids that are protonated at a pH of 5.5 or less are not protonated at a pH of 7.4.

Numbered embodiment 105. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-100, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5 or less, and wherein the one or more amino acids that are protonated at a pH of 5 or less are not protonated at a pH of 7.4.

Numbered embodiment 106. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-105, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polynucleotide is present within an endosome.

Numbered embodiment 107. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-106, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polynucleotide is present within a lysosome.

Numbered embodiment 108. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-107, wherein the pH responsive polypeptide domain comprises an amino acid sequence

having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 336-349.

Numbered embodiment 109. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 110. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 111. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 112. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 113. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 114. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 115. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 116. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 6 or less the endosomal escape peptide has an affinity to the pH

responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 117. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 118. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-108, wherein at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

Numbered embodiment 119. The composition of any one of numbered embodiments 81, 84, and 90-118, wherein the endosomal escape peptide is released from the protein complex when the protein complex is present within an endosome.

Numbered embodiment 120. The composition of any one of numbered embodiments 83, 86, and 92-119, wherein the endosomal escape peptide is released from the protein complex when the protein complex is present within a lysosome.

Numbered embodiment 121. The composition of any one of numbered embodiments 83, 86, and 92-120, wherein the endosomal escape peptide comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 152-335.

Numbered embodiment 122. The composition of any one of numbered embodiments 82, 83, 85-87, and 94-121, wherein the pH responsive polypeptide domain is capable of binding to two or more endosomal escape peptides at pH 7.4.

Numbered embodiment 123. The composition of any one of numbered embodiments 92-122, wherein the protein complex comprises two or more copies of the endosomal escape peptide.

Numbered embodiment 124. The composition of any one of numbered embodiments 79-123, wherein the fusion polynucleotide further comprises a linker linking the polynucleotide and one or more other peptide sequences of the fusion polynucleotide.

Numbered embodiment 125. The composition of numbered embodiment 124, wherein the linker is a proteolytically cleavable linker.

Numbered embodiment 126. The composition of any one of numbered embodiments 79-125, wherein the fusion polynucleotide further comprises one or more nuclear localization sequences (NLSs).

Numbered embodiment 127. The composition of numbered embodiment 126, wherein the one or more NLSs comprise the amino acid sequence K(K/R)X(K/R), and where X is any amino acid.

Numbered embodiment 128. The composition of numbered embodiment 126, wherein the one or more NLSs comprise the amino acid sequence PKKKRKV.

Numbered embodiment 129. The composition of any one of numbered embodiments 79-128, wherein the one or more NLSs are at the 5′ end of the polynucleotide.

Numbered embodiment 130. The composition of any one of numbered embodiments 79-129, wherein the polynucleotide is RNA.

Numbered embodiment 131. The composition of any one of numbered embodiments 79-130, wherein the fusion polynucleotide further comprises at least one additional heterologous polypeptide.

Numbered embodiment 132. A nucleic acid comprising a nucleotide sequence encoding the fusion polynucleotide of the composition of any one of numbered embodiments 1-131.

Numbered embodiment 133. The nucleic acid of numbered embodiment 132, wherein the nucleotide sequence is operably linked to a transcriptional control element, optionally wherein the transcriptional control element is a promoter.

Numbered embodiment 134. A recombinant expression vector comprising the nucleic acid of numbered embodiment 133.

Numbered embodiment 135. A cell comprising the composition of any one of numbered embodiments 1-131 or the nucleic acid of numbered embodiment 132 or 133, or the recombinant expression vector of numbered embodiment 134.

Numbered embodiment 136. The cell of numbered embodiment 135, wherein the cell is a eukaryotic cell.

Numbered embodiment 137. The cell of numbered embodiment 135 or 136, wherein the cell is in vitro.

Numbered embodiment 138. The cell of numbered embodiment 135 or 136, wherein the cell is in vivo.

Numbered embodiment 139. A method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with:

(a) the composition of any one of numbered embodiments 79-131; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

Numbered embodiment 140. A method of modifying a target nucleic acid in a eukaryotic cell, the method comprising introducing into the eukaryotic cell:

(a) the composition of any one of numbered embodiments 79-131; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

Numbered embodiment 141. The method of numbered embodiment 140, further comprising introducing into the cell a donor nucleic acid.

Numbered embodiment 142. The method of numbered embodiment 140 or 141, wherein the cell is in vitro.

Numbered embodiment 143. The method of numbered embodiment 140 or 141, wherein the cell is in vivo.

Numbered embodiment 144. The method of any one of numbered embodiments 140-143, wherein the cell is a mammalian cell, an insect cell, an avian cell, a reptile cell, an amphibian cell, an arachnid cell, a protozoan cell, or a plant cell.

Numbered embodiment 145. The method of any one of numbered embodiments 140-144, wherein the target nucleic acid is selected from: double stranded DNA, single stranded DNA, RNA, genomic DNA, and extrachromosomal DNA.

Numbered embodiment 146. The method of any one of numbered embodiments 140-145, wherein the modifying comprises genome editing.

EXAMPLES

[00301] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric. Standard abbreviations may be used, e.g., bp, base pair(s); kb, kilobase(s); pl, picoliter(s); s or sec, second(s); min, minute(s); h or hr, hour(s); aa, amino acid(s); kb, kilobase(s); bp, base pair(s); nt, nucleotide(s); i.m., intramuscular(ly); i.p., intraperitoneal(ly); s.c., subcutaneous(ly); and the like.

Example 1: Developing and testing endosomal escape peptide

[00302] To enable cytoplasmic delivery of receptor-targeted Cas9-based RNPs, peptides for endosomal escape were developed. These peptides contain pH-responsive domains (i.e. histidines and glutamic acids) that become protonated at endosomal/lysosomal pH (pH 5 to 6.6) and cause the peptides to disrupt the endosomal membrane. To expedite discovery of functionally competent endosomal escape peptide sequences, a pH-responsive red blood cell (RBC) hemolysis screen was devised as described herein. In this screen, RBCs are purified from whole blood and resuspended in various pH buffers (pH 5.0 to 7.4) and incubated with crude peptides, synthesized in a 96-well plate format. EEPs that have little to no activity at pH 7.4, with increasing hemolysis at lower pH, were selected for and further optimized. In addition, these peptides were ordered with HPLC-purification and combined in trans with RNPs and tested for editing on cells in vitro. Provided herein below (see for example **Table 4**) is a list of peptide hits from three hemolysis screens.

Table 4. Amino acid sequences of exemplary Endosomal Escape Peptides

SEQ ID NO.	Peptide name	Amino acid sequence
152	app5.3	GPSQPTYPGDDAPVRDLIRFYRDLRRYLNVVTR
153	FHV-TA (39-49)	NRARRNRRRVR
154	FHV (40-49)	RTRRNRRRVR
155	CM18-PTD4	KWKLFKKIGAVLKVLTTGYARAAARQARA
156	INF7	GLFEAIEGFIENGWEGMIDGWYG
157	A22p.4	HTPGNSNHWHHLQENHHGRPRR
158	GALA	WEAALAEALAEALAEHLAEALAEALEALAA
159	TH	AGYLLGHINLHHLAHLLLIL
160	H5WYG	GLFHAIAHFIHGGWHGLIHGWYG
161	A22p	HTPGNSNKWKHLQENKKGRPRR
162	peptide_623	KILRGVSKKIMRTFLRRILTGKK
163	14980	SKVWRHWRRFWHRAHRKL
164	12191	SPHRPRHSRLQREPQVQWLEQQVAKRRTKR
165	766	RKSKEKIGKEFKRIVQRIKDFLRNLVPRTES
166	cpp_A8	HRLRHALAHLLHKLKHLLHALAHRLRH
167	12202	YYHFWHRGVTKRSLSPHRPRHSR
168	peptide_pm_160	KWKSFLKTFKSAAKTVLHTALKAISS
169	765	GDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES
170	peptide_328	ISLNPPRSTIAMRAINNYRWRSKNQNTFLR
171	peptide_388	RRRCPIVVIRVCRR
172	peptide_pm_132	RGGRLCYCRRRFCVCVGR
173	peptide_pm_128	QCRRLCYKQRCVTYCRGR
174	16603	KLIHRLKTVFKKVWHFLGHL
175	16595	RVFKKFHPIYHRVWRLWGHL

176	peptide_pm_181	FKRLEKLFKKIWNWK
177	17458	SHHRHFRSHQCRRQRSNSCDR
178	cpp_RLW	KHKLLHLLHLLALLWLHLLHLLKHK
179	cpp_TP10	GWTLNSAGYLLGKLKALAALAKKIL
180	17467	SHHRCRRSHR
181	9991	NLVSDIIGSKKHMEKLISIIKKCR
182	cpp_No.440	IPSRWKDQFWKRWHY
183	peptide_278	FLPLLAGVVANFLPQIICKIARK
184	peptide_pm_316	GFWGKLFKLGLHGIGLLHLHL
185	peptide_pm_367	FVQWFSRFLGRIL
186	3916	GLLSVLGSVVKHVIPHVVPVIAEHL
187	peptide_pm_226	SWLRDIWDWACEVLSDFK
188	peptide_pm_129	SWLRDIWDWECEVLSDFK
189	Unknown	GFNEIVQDIEDFLQNLV
190	A5k	GLFEKIEGFIENGWEGMIDGWYGYGRKKRRQRR
191	INF7_TAT-4RH	GLFEAIEGFIENGWEGMIDGWYGYHKKHHQHR
192	app5.3_D	GPSQPTYPGEEAPVRELIRFYRELRRYLNVVTR
193	a22p.4.1	HTPGNSNHWHHLQDNHHGRPRR
194	a22p.4.3	HTPGNSNHWHHLQWNHHGRPRR
195	a22p.4.4	HTPGNSNRWHHLQENHHGRPRR
196	KALA_H.1	WEAHLAHALAHALAHHLAHALAHALHACEA
197	KALA_H.2	WEAKLAKALAHALAHHLAHALAKALKACEA
198	EB-1_H.1	LIRLWSHLIHIWFQNHHLKWKKK
199	EB-1_H.2	LIRLWSHLIHIWFQNHHLHWKKK
200	EB-1_H.3	LIRLWSHLIHIWFQNHHLHWHHH
201	9592.his_1	IWSFLIHAATHLLPSLFGGGHHDS
202	9592.his_2	IWSFLIHAATHLLPSLFGGGKHDS
203	peptide_pm_332.his_2	ILPWHLPWWPWHR
204	17985.his_2	GWYEIIHKIYHWLK
205	LAH4	KKALLALALHHLAHLALHLALALKKA
206	IMT-P8	RRWRRWNRFNRRRCR
207	S18	LLKLWSRLLKLWTQGGSGGGSGRRLKAKRAKA
208	S85	LLKIWSRLIKIWTQGRRLGARAQAR
209	IMT-P8_his.3	RRWRRWNHFNHHHCH
210	Pep115	GWFHALLHLLHSLWHLLLHA
211	Pep086 GLFH(3-Me)	ALLHLLHSLWHLLLHA
212	Pep030	GLFHALLHLLRSLWHLLLHA
213	Pep015	GLWHALLHLLHSLWHLLLHA
214	Pep150	GLFHALLHLLHSLWHMLLHA
215	Pep013	GLLHALLHLLHSLWHLLLHA
216	Pep005	GLFHALLHLLHLLWHLLLHA

217	Pep028	GLFRALLHLLHSLWHLLLHA
218	Pep035	GLFHALLSLLSSLWHLLLHA
219	Pep029	GLFHALLRLLHSLWHLLLHA
220	Pep023	GLFHAFLHLLHSLWHLLLHA
221	Pep055	GLFHALLHLLHGLWHLLLHA
222	Pep142	GLFHALLHLLHSLWHLMLHA
223	ppTG21	GLFHALLHLLHSLWHLLLHA
224	Pep177	GLFHALLHLLHSLWHLLMHA
225	Pep006	GLFHALLHLLHALWHLLLHA
226	Pep049	WLFHALLHLLHSLWHLLLHA
227	Pep053	GLFHALLHLLHTLWHLLLHA
228	Pep034	GLFHALLSLLHSLWHLLLHA
229	Pep007	GLFHALLHLLHSLWHWLLHA
230	Pep027	GLFHALLHLLHSLWHLLLRA
231	Pep148	VLFHALLHLLHSLWHLLLHA
232	Pep051	GLFHALLHLLHQLWHLLLHA
233	Pep008	GLFHALLHLLHSLWHLFLHA
234	Pep122	GCFHALLHLLHSLWHLLLHA
235	Pep116	GVFHALLHLLHSLWHLLLHA
236	Pep158	GLFHALLHLLHSLWHLILHA
237	Pep048	YLFHALLHLLHSLWHLLLHA
238	Pep031	GLFHALLHLLHSLWRLLLHA
239	Pep004	GLFHALLHLLHSLWHLLLHAGGSGGSGLFHALLHLLHSLWHLLLHA
240	Pep036	GLFHALLSLLHSLWSLLLHA
241	Pep188	GLFHCLLHLLHSLWHLLLHA
242	Pep100	GLFHALLHLWHSLWHLLLHA
243	Pep052	GLFHALLHLLHNLWHLLLHA
244	Pep081	GLFHALLHLLHSLWH(1-Me)LLLH(1-Me)A
245	Pep152	GLFHALLHLLHSVWHLLLHA
246	Pep065	GLFHALLELLHSLWHLLLHA
247	Pep050	GHLFHALLHLLHSLWHLLLHA
248	Pep119	GLFHGLLHLLHSLWHLLLHA
249	Pep087	GLFHALLH(3-Me)LLHSLWHLLLHA
250	Pep187	GLFHALIHLLHSLWHLLLHA
251	Pep056	GLFHALLHLLHDLWHLLLHA
252	Pep171	GLFHAILHLLHSLWHLLLHA
253	Pep026	GLFGALLGLLHSLWHLLLHA
254	Pep165	GLFHALLHLLHSLWHLLLMA
255	Pep032	GLFHALLHLLHSLWRLLLRA
256	Pep016	GLFHALLHLLHHLWHLLLHA
257	Pep009	GFFHALLHLLHSLWHLLLHA
258	Pep153	GLFHALCHLLHSLWHLLLHA
259	Pep136	GLFHALLHLLHSLWHLLLCA
260	Pep067	GLFHALLHLLHSLWELLLHA
261	Pep155	GLFHALLHLVHSLWHLLLHA
262	Pep014	GLYHALLHLLHSLWHLLLHA
263	Pep022	GLFHALLFLLHSLWHLLLHA
264	Pep095	GLFHALWHLLHSLWHLLLHA

265	Pep039	GLFHALLHLLHSLFHLLLHA
266	Pep047	LLFHALLHLLHSLWHLLLHA
267	Pep161	GLFHALLHLCHSLWHLLLHA
268	Pep191	SVGSGLFHALLHLLHSLWHLLLHA
269	Pep104	GAFHALLHLLHSLWHLLLHA
270	Pep185	GLFHQLLHLLHSLWHLLLHA
271	Pep042	GLFHALLHLLHSLAHLLLHA
272	Pep080	GLFHALLHLLHSLWHLLLH(1-Me)A
273	Pep110	GLFHALLHLLHSLWHALLHA
274	Pep163	GLFHALLHLLHSLWHLLLTA
275	Pep025	GLFGALLHLLHSLWHLLLHA
276	Pep040	GLFHALLHLLHSLLHLLLHA
277	Pep144	GLTHALLHLLHSLWHLLLHA
278	Pep057	GLFHALLHLLHELWHLLLHA
279	Pep079	GLFHALLHLLHSLWH(1-Me)LLLHA
280	Pep068	GLFHALLHLLHSLWHLLLEA
281	Pep106	GLFHALAHLLHSLWHLLLHA
282	Pep126	GLFHALLHLLHSLWHLLRHA
283	Pep189	GLFHALVHLLHSLWHLLLHA
284	Pep078	GLFHALLHLLH(1-Me)SLWHLLLHA
285	Pep066	GLFHALLHLLESLWHLLLHA
286	Pep071	GLFHALLDLLHSLWHLLLHA
287	Pep037	GLFHALLSLLHSLWHLLLSA
288	Pep131	RLFHALLHLLHSLWHLLLHA
289	Pep124	GLFHALLHLLHSLWHLLLFA
290	Pep145	GLFHALLHLIHSLWHLLLHA
291	Pep092	GLFH(3-Me)ALLH(3-Me)LLHSLWHLLLHA
292	Pep046	HLFHALLHLLHSLWHLLLHA
293	Pep146	GLFHYLLHLLHSLWHLLLHA
294	Pep096	GLFHALLHWLHSLWHLLLHA
295	Pep162	GLFHALLHLLHSLWHELLHA
296	Pep077	GLFHALLH(1-Me)LLHSLWHLLLHA
297	Pep102	GLFHALAHLLHSLWHLLAHA
298	Pep135	GLHHALLHLLHSLWHLLLHA
299	Pep033	GLFSALLHLLHSLWHLLLHA
300	Pep129	GGFHALLHLLHSLWHLLLHA
301	Pep107	GLFHALLHALHSLWHLLLHA
302	Pep105	GLFHAALHLLHSLWHLLLHA
303	Pep192	GSGLFHALLHLLHSLWHLLLHA
304	Pep179	GLNHALLHLLHSLWHLLLHA
305	Pep130	GLFHALLHYLHSLWHLLLHA
306	Pep072	GLFHALLHLLDSLWHLLLHA
307	Pep182	MLFHALLHLLHSLWHLLLHA
308	Pep076	GLFH(1-Me)ALLHLLHSLWHLLLHA
309	Pep180	GLFHAYLHLLHSLWHLLLHA
310	Pep038	GLFHALLHLLHSLYHLLLHA
311	Pep089	GLFHALLHLLHSLWH(3-Me)LLLHA
312	Pep064	GLFEALLHLLHSLWHLLLHA

313	Pep024	GLFFALLFLLHSLWHLLLHA
314	Pep099	GLFHALLHFLHSLWHLLLHA
315	Pep074	GLFHALLHLLHSLWHLLLDA
316	Pep127	GLFHALLHLLHSLWHLLGHA
317	Pep134	GLFHALLHLLHSLWHQLLHA
318	Pep178	GLFHALLHLLHSLWHLLIHA
319	Pep172	GLQHALLHLLHSLWHLLLHA
320	Pep109	GLFHALLHLLHSAWHLLLHA
321	Pep128	GLFHAWLHLLHSLWHLLLHA
322	Pep073	GLFHALLHLLHSLWDLLLHA
323	Pep176	GLFHALLHLLHSLSHLLLHA
324	Pep070	GLFDALLHLLHSLWHLLLHA
325	Pep157	GLFHALNHLLHSLWHLLLHA
326	Pep168	GLFHALLHLLHSLWHLTLHA
327	Pep002	GLFHALLHLLHALLHLLHSLWHLLLHA
328	Pep125	GLEHALLHLLHSLWHLLLHA
329	Pep159	GLFHALLHLLHSLWHLLPHA
330	Pep186	GEFHALLHLLHSLWHLLLHA
331	Pep111	GLFHALLHLLHSLWHLALHA
332	Pep108	GLFHALLHLAHSLWHLLLHA
333	Pep160	GLFHALLHLLHSLKHLLLHA
334	Pep069	GLFDALLDLLDSLWDLLLDA
335	Pep059	GLFHALLELLESLWELLLEA

Note: H(1-Me) is 1-methyl-histidine and H(3-Me) is 3-methyl-histidine

Example 2: Generation of receptor targeted mini-proteins

[00303] For in vivo editing with RNPs, receptor targeting can allow for delivery to specific tissues and cell-types. To enable receptor-binding and uptake, computationally designed mini-proteins (e.g., targeting moiety) were made and validated in cells and mice. The benefit of mini-proteins, unlike antibodies, includes, but is not limited to, their ability to be expressed in bacteria. Because Cas9 is also expressed in bacteria, mini-proteins can be fused to Cas9 to express a single fusion protein. Initially, hydrophobic binding domains were identified on specific receptors, and scaffolds are screened against these domains, followed by various computational techniques to optimize amino acid sequence to enable high-affinity to the target receptor. In the case of TfR1, a previously designed mini-protein, incorporating alpha-helices and beta sheets extending of the TfR1 beta sheet was utilized as the initial scaffold. New topology and interface design was applied to the scaffold and a library of 1500 designs was made and expressed in yeast. Yeast display, expressing the scaffolds was screened with the TfR1 receptor at various concentrations down to 1 nM. Yeast were sequenced by next-generation sequencing (NGS) and the list of scaffolds with high enrichment was generated (see for example **Table 5** for list of mini-proteins tested). Scaffolds were then expressed and purified in bacteria, followed by measuring uptake in HEK293 cells overexpressing hTfR1. The scaffolds that

showed high uptake in cells at low concentrations were then fused to SpyCas9 at the N-terminus and retested for uptake and editing in cells.

Table 5. Amino acid sequences of exemplary mini-proteins (e.g., targeting moiety) targeting TfR1

SEQ ID NO.	Amino acid sequence	Mini-protein
31	MVVYSDDPAVALEIATKYGGKPWEEIEKAAKEAAKKGVAVVVVVGTDEEE TIAYNTALANDTGTGFEQAKKIADAIKAAVPNAKVVLA	L271
32	MAEEVYKLVEEMLEKGETDVKKIAKAVKEKFPEVNVIEVGNVAIFIVGGTE GLEERKAAAEKVYKEGIAAGWNTTELLTALLLATGADVVVA	L1179
33	TADEVYKLVEEIGGSGPEAAERAAEAVEKKFPDVYVVRLPDGTAVFVVGHE GKEEKQAIALAVYEANKAAGITDPIERKINYLTAIPGAAVVYG	L647
34	AAAAAAARAAFEATLERLKADGTYDDPIAAGTALLLAVPGARVVVYDKKY AVFITDEKADEVYKLAEELLAKGATVEEISKAIKEAVPTVTVLLG	L56
35	GAALEEARRLAREVVERAKREGLDPLTTALRLLEAIPGASVVVAAESYAVFI TKEKAEEVYKLAEEFDPKSAEDIRELAELVREKFPEVEVVLA	L41
36	GGEEAYELAEKLLAKGATPEEIAEAVAKELAGRVIVVRVGDVVVLVEGDEA EAPRRQALAEEVYERSRTLSPQERLEAFLTGVPGARVVVG	L1139
37	AAARAAARAAFEGVLAQAKAEGITDPLEIALRLLEAIPDAAVVYANPYTVF VSREHADEVYRLVEKAIEEGITDSREIAKRVREAVPDVEVLLG	L1442
38	EEERKELEEKVKKVLEEAKKEGITDPTEIAERILLETGARVVVAGEKYAAFF APPEYAEEVYKLAEEALKKVKDVEEVARIVREKVPEVFVVLA	L1143
39	AAEAAAKEAAARAVVEAGKKEGKSPLETAIALLEAVPGVRVVYYRERFAA FFAPEEDAPKVYKLVEEASKKYENEEEVAAYVKEKVPTVEVVIL	L1107
40	AADEVYKLVDELLEKGETNVEVVAKAVREKFPEVTVITVGDVAVFIVGKEE EMEEREKKAKAVYEEAKKEGIDRETLLIRMLEATGADVVVG	L628
41	EEKWKEAEERFRKTLEKMKEEGVTDPAEIATALLLSIPDASVVVATDKYAV FIAPKEDHDKVFELAEKLLEEGGTDVADIAKKVQEAIPSVKVILG	L313
42	AADAVFAVAEELAGKASVEEIAEAIRERVEGNVLLIVKDDVLVAVVGEIAD PEAVRARFDEVYEASKDLDPQSRLTAILLATGADVVVG	L53
43	LADKVYEIAESLLEFDDVKAIAEAVEKAVPGVHVVVIDDTAVFTVGGNKEE AEKKAKEVEEELKAKGITDKEERNLAFLEAIPNAKVVIS	L194
44	AAAREAAEAAVRAVLARAKAEGTSLEETNIALLEAVPDAAVVVGYERYAV FITKKDAEGVYKLAEENLPSSEEGIRELAERVREAFPDVHVIVD	L134
45	GMDKVYEIADKLIDEVENIEEIAKKIEEEVPGVRVIIIGNYAVAIEGGEENAEE YEAKAKKVIDEAKAKGITDPNEIGIALLLETGAGVVVS	L152
46	AADKVYAIAESSQGITDPREMAELVSAEVGGEVEVLVLDDFAVFVVGGSED REEIEERLKEVVERGKKEGTPKEELAIELLLESGGAVVFT	L147
47	DAALEKAYKAASEALDKFGDNDMKGIAKYIADTVGDAVEVLYHKNVVYV KGGVVVSNNPAVALYLAENLAGAPKEEVFAAIDKLE	L1322
48	MGEEVYKIAESALGKMSPEEVADLVEEKIGGKVRAIRLPGGYVVFVEGNDG FEEAEARARAAWEEIVAEGLTDDPTAVLTRLLLATGGSVAIA	L1330
49	MEKEIIEEVKELLKTKTKEETAVALVEKFPEIKVVVVTDKIVAVVADDELGE KAYKAAEEALEKFPKDAAEEVAKYIKEKVGEGVKVISI	L48
50	MVVVSEDPELALKLVEELANGTKEEVEKFVKEYSGGGVKVVVVLTNDPET AEKAYKAAEEALDKFGINNTEEIAKYIKEKVGEDAEVVNI	L1130
51	DELYEEALKLGEKIVEEGKKNGWSFEEIGERLLTAVKGAALVYVTEEKVVI VPEEVADEVYEAVSKMLEEGNFDEEEIIEKIKELAKGKAKVVVV	L1298
52	GEKIEEFKKKFKAAMEENKNNPTTTPAEIALEAYPEAGLVIAYGDLVLLVHK EVADEAYKIAEEILGDDEESVVKAYEILKEKSGGKVTLVKI	L509
53	MAEKIYEIAEKNLEGHSVAEIGELIEKEVPGARTIVVGDVLAVVVGGTEDDA TLRARAEAVVARGKAEGLSRLETLERILTEVPNAKVVVG	L21
54	EEAIEKLKEKIKEIVKEAKEKGLNTEETALRILTEVGAKLVVLTDGYVVIVLD PEDADKAYELAEELLKKYPNDSEIIVNKLKEAFPNVIVVEV	L1106
55	ALDKVYEIAEKTLGKHNVKEVAEIVKKEVPEVEVIIVGDIAIFVVGGDPAEK RAAAEAVLERAKAEGLNRLETLLAFLEAVPGAGVVVG	L52

56	EELRKRVEAAVEEVWKQKENMNKEELATALLLKSGASVVVAAERYAVIVV PKEKAEEAYKIAEENLPENDEELEELAERIEEEVPGAVVVLT	L1293
57	SAELAAKRAAFEAALAQSKSLNTTDALTLLLLATGASVGVWAGEYAVFLV PKEKEEEVYEIAEEMLKEGLKDVRELAERVEREVPGVVVLLG	L623
58	MEELEREERIVRELVEEGKREGIDKLTLAEKLLEAVPRARVVVVDEEYAVFI ADPSIADKVYEIAEKALEEGVKDVREVAERVKKEVPGVHVIVA	L1111
59	MELDEKAYEAAREALGTGENRGKKMAEYIKEKVPEVKVVLVGDDLVIVFR GDADVEAVVATAEKLLKEVSKEQAALELVEKFPEIIVVVV	L572
60	EEERERLRRLVEEVLERGKKEGVSPLELAEKLLLEIPGARVVAYDEKFAVFI ADDETGEKVYKLAEEKLGKVSVEELAELVRKEFPSVEVLLA	L279
61	EEEFEKARAIFDAIVEEAKAKGWDAVTTGIALLEGVPGARVVAYTDRWAVF IADPKVAEEVYKLAEEAMEEVETVEEIAEVVKEKVPEVKVLLS	L1159
62	DPLVEAALEAARIARERRAEDGGDDWVVVVIKDPEAAAEVAEAVGAEEVR QIGDYTVLFARGVSLEEAVAAAEAASRKYNSPYAVIGL	L365
63	MSEEQKAYKVAEKELFGDSEENAKRIGEAISKETGLKVIVHGKVVVVLKGE VDEAAVKAAIDALEGKSDEDRATALLQQFPEIVLVVV	L394
64	VHVIVIEPDENTVKIAEKVKEEAGNDPEEIYKVAEKLVDEYPTTGFAVFITDP TGAAEKEATARAIYAEAKAKGVDRLTTLERLLLGTGAAVVVG	L45
65	GMDEELATQAVEIAAQVGAKKAYAIVGDGFILVIIVGKGAAAKKAAEKILE LAKKAGIKGEVYMVEGGDDEELIEELLKKWEEIKAKLGK	L127
66	AAAQARADADVIVAHFGDLPERERTVLRLAAIAAAAGAEKVEAWEEGGRY HVRITGTDAAAKAGAAAARAWAAERGLDVEVTLG	L238
67	MGDKVYEVAEKLLEGSSPEKIAEAVKKAVPGVQVFVEGDFVAFVVGGEED REEREARVREIVERAKREGWSTEETLLAILEGVPGATVVVG	L1435
68	AYVVVVVSDGDLTPAAAAAAAEGATVRVEGNIVIAEVPDEATAMAVLTNA AAAVGDARGFIIPAGEEGLEERIREEAERLRKREEEVRNGP	L135
69	MEAAEKAYEVAEKYFGTGEENAKKIAEAIKKEAGDELDVIVVKDIVIVGKG FDKEEVKKRVEELLKEKSLEETALELVEEFPKIYVVAV	L1140
70	DAEFEAAEAAAKKVIEEAKKKGLIGKDPITVMTDLLLAIDGASTVVYDKEY AVFIHKEKADEVYKLADELVGKYNVKEIAEIVKKEIPGVRVVML	L1294
71	MVVVSDNPEAALAIAETNTGNMEEAKKMAKELTGGKGWIVVAFADDEKA EKAYKVAEKAIEIGDNKENAKKIAEAIRKEVPGVEVIEL	L703
72	MELEERVRKIVKEVVEKGKKEGKSPEEIAIAILEAVPEASVVYYSEDFVVLFP HKVAEEAYKLAEKLLEELTDEEEIKKKLLEEFPEAILVEL	L1108
73	EKELEEKREKAKEIIEKGIKEGKSSVEIAEELLKEVGASVVYINDGVAVFLAP PEVADKAYKIVESFGGKKSAEEVAKAVVEETGGKVEVFLV	L31
74	AAAQARADADVIVAHFGDLPERERTVLRLAAIAAAAGAEKVEAWEEGGRY HVRITGTDEAAKEGAEAARAWAEERGLDVEVTLG	L991
75	EEKRKEAIEKIKEVIEKYKKTGDLEALGIAVIEKSGASVVVSLPNYVVAVLDP KLADEVYKLAEELIEKTDDEEEIAKAIKEAFPEVEVIKL	L1397
76	EEERRAREARARELLAAARAAGIWESDPQALGTALLTGVPGARVVVVAER VAVFVTDEKADEVYELAEKNLPTSAEGVEELAELVRKELPDVEVILL	L1504
77	LEEKVYELAEEALKTLTDVKEIAEYVKEKVGDKVTVFIIGDIAVFIVGDLSEE ELEEKKKKAEEIYKQPLSTTDKMIEILKATGASVVVA	L347
78	MELLEKVYEVARKAIGKGENEGRVIAEAIREETGADVLLIGDGLVVVLTGN QDKEAIKEYYEKNKGDDEATAVEIAEKFGAGVVVV	L51
79	DEAAEKAYEAAEEALEKYPGREDGKKRADYIKEAVGDELDVRLFHPEVVV VGRGFDEEAVRAYVEKNLGDLEKLALELIEKEPKIVVVVV	L1102
80	AARRVYEIVERAGGSGEEEAERVAEIVKKLVPDVVIFVHEDGFVVFVVNPE GAEEQRRRLEEVKEEAKRRGITDPQEIAVAYLEAVPGAAVVFR	L813
81	SELEQGVYELAEKLYGEGKENMERIAKAIKEKFPEFTVILVEDIVVVFRNAD EEEIKEAVKRLLKEKSKEETALELVEKFPGVVVVVA	L968
82	DEEEIQKAIEELLKEKSLEEAAIEIAQRFGAAVVVVTKTLVVIKDDVILFGEG DREGKHIAEYIRRELEGFNGSKHERSTKAWEAAQKAF	L571
83	AARRVYEIVERANGGGAEEARRVADIVRKLVPEVEVLIVDPEFVVFVVGRE GAEEQRRRLEEVRERARREGLSRQELAVARLEAIPKAAVVFG	L1185
84	MADEVYRLAEEALGLTDVREKAAAVLARVGGRVRAFVFDDIAVFVDGGED RWAEIRARLEAVWRAIKARGGALTTEEKLTAFLEAVPDAVVAVA	L748

85	AVVVSENPEAALELVETVQDDPEKARALAKELTGGKKPIVVAFATEPELGQ KVYDVAKEAFPNYGGEEGAKAIADAIKDTLGDEVTVIAF	L381
86	AAREVYEIVERALERGEEAERVAEIVKKLVGGEVTIFVHGDLVVFVVGREG EEEQRRRLEEVKEEAKRGLSPQEAAVLALEAVGAAVVYV	L537
87	MAAEEAAYKEAEKALEKGSGPELGRYAQEYLSKNVGGELEVHLGHPDIVV VHRGEDPAEIDAAFEELAHLSLEDRGIEIARRFPGVVVVVA	L287
88	VEVILIEPDEEIKKIAEEVKKEVPGGDPELIYKVAEKLTEKYPFKGTAVFVAG GLATVVSGGDTELATKLTLAAQGNAEQREANVKKVVEEYKK	L1262
89	MVVVSEGDTETSIKLLLATQGDKESNEAAFKKVYAEAKAKGKGRFAVLVA PESVADEVYKVAEEYLNTEDLEEVAEKVKEAVPEVEVFLD	L277
90	SLFEALEKAAKLLKERSGSDDDVLVVIKNPEAAEDAAKAINAEEVTQIGDW TVLFARGVSFDEAVAAAEAASQKYNTPFAVLKL	L95
91	GVEEAYKIAEEALKGGNPKEIAKKIKEGVGDTIEVFDVGGEFAVVVPKGAG DVEARKAAAEKAYKAAKAAGETGAKAAIAIAEAVGVPSVVVA	L951
92	EEEFKEKEKIAKEVVEKGKKNGTPKEDILVDLLLATNARVVVSDPYVAFLA DEENGEKVYEIAEKLVGKKNGKEIAEVVKKEVPNVKVIVL	L772
93	MEEIKKKEEKVEEVLERAKKEGITDPTEIALELLEATGASVVVYDEKYAVFI DKENAEEVYKIAEELLGGKTVEEIAEEVKKKIPNVKVFLK	L811
94	AAEKVYEIAERELGGGREDARRVAEIVRREVPDVVVVEVGDYVLFVVGRH ADAERLRAIAEEALARVKAEGLENDPEAAGLLFLEAVPDAAVVVG	L992
95	GGDEAYKIALEAGENASVEEVAERVRRELEGRVIVLRFGDIAVFVDAEAGD EETHRERLEAAWEEALKIEDPVERAELFLLAVGASVVFR	L1112
96	DVVVADDEEAALKLLEAAQGDEETIRKNVAEVAKELEGKGVKFVVFVADP EVADQVYEIADETIKKVKDPEEIAEAVKAKIGGKVLVIVE	L1469
97	MVVVSFGDPQAALELVEKYAGAPLEEVKKAIKELGSEAIVYVYAEDDEAIE KAYEVAEKYFGGSGLEEAKRVAKKIEEEVGGKVEVVAE	L532
98	AARRVYEIVERANLGPEEAKRVADIVKKLVPDVIIFVHDGFVVFVVGREGEE EQRRRLEEVKEEAKRIGDPQELAVLYLEAVPGAAVVFR	L708
99	MVEEIKKAAKELSSLPAEEAAIELTEKFGLAVVVVDRLVIVTANSDELSTKA YEIAREVYGGGEEDSRRIAEAIKEALGPDVEVVIL	L217
100	DEREEEARRRVEEVKERAKREGLSRQELAVALLEAVPDAAVVVVTEEAVV FIVKDAARRVYEIAERALPSSREELRRVADIVKKLVPDVIIIFH	L151
101	AAEEVYKIAEEVGGDKSAKEVAKAVKEAIGGKVYIIEFDDDTVVFFAGGGV VYGNNLEVNLKLLESIGKNKEETEKNVKETYEKIKN	L553
102	AAREVYEIVERANGGPAEEAERVAEIVKKLVPEVRIIRVAPDFVVFVVGDEG EEELRRRVEEVEERARREGWSLQERAVALLEAVPQAAVVFG	L607
103	MVVLTDNPEVGLELVEKLSQGTPEEVKAFIDEKLGGKGVVVYAVAETEEM QEKAYEVAEAAYRPLDAESAKEIAARIEEEVGEGVTVYSL	L705
104	EEEFKEAEKKAKEVIKKAKEEGVGETELLIRLLEAIPNARVVVANGYAVFLA EPEYADEVYKLAEKYGEKNNAEEVAEVVEKELKGKVRVIKV	L569
105	AAEKVYEIAEKEVVEDEEGIKELAEKVKEEVEGEVLVFVKGNIAVFVVGDF ENKEEIEKKLEEAEKEIKEKGLSQTEALLTFLEKVNADVVYG	L485
106	MEMDERVYEVAEEALEASGGPEGAEAVARRVAEAVPGVKTIAVGTIVAIGR DFDEAAVRAAIEALAKTLSPEEAAEALIQAFPQITVVVV	L282
107	MVVFSVENVEFAEEVVENFSNKPMEEIEKYLKEKGAKDTVVAKATDPAKE EKAYEIAKKYFGSGPGEAEKIAEAIKEEVGEEDVLLV	L1336
108	AARRVYEIVERARGEDAEEAKRVADIVKKLVPEVIIFVHDDTVVFVVGGNE EEQRRRLEEVKERAKREGWNRQDRAVLYLEAVPTAAVVYT	L812
109	MTEKVYELAEKEVGKDSIEEIAEAIKKAVPGVHVLLFDGKYAVFFSGGGVV VTEGDTELAIKLLEAIGSTVEETEKNVKKVIEAAKK	L280
110	MVVVSMSGTEAALQLVEEFPNGDFEEIKKKAKELAKGQPLVVVRADDELG EKVYEAAREAFGGGDAEKAKEIGEYIKEKTGAEVIVL	L328
111	DALEKKVWEVASKALDLGSGPEAMKRIAAEIAAKTGAEVVAHGDIVYVSG GNVVVALGGEEAALQLVEEVAGKPLEEVLAKMRELA	L333
112	MEEELKVYEVALKALEEAGENRGKQIAEYIKENAGDKFKVELISPDLVVVT DGDVDLEGIKKLYEEKKGDLEAAGVAIAEKFPGVVVVVV	L941
113	MVVVTDGDTEAALALLNAIGGTPEETRAAVEAALAEAKAKGVKAAVFHTG KNADEVYKLAEELLEVDDIEEIAEAVREKVPGVTVITL	L925

114	MARRVYEIVERARGKDEAKRVADIVKKLVPEVQIFVFGDVVVFVVGDPATE EEQRRRLEEVMEEGKREGLDKQTLAVRCLEAVNAAVVWV	L1266
115	MVVVAVEGKEAALKLAEKYPNGDPEEIKKEAKKLTGGKPVVVAYGGTTNE EQLKIYKAAEEALEKNGDKEGAKEIAKYLKENGAKEVIVL	L579
116	MKEEVKKLAEELKDISEEEIAEELLRKFPEIVVVVANPLVVVFADDETGEKA YKAAEKAIGKGENEGENIAKYIKEEAGDDVEVILL	L960
117	MADEKVAVEAVEIAAQVGVKEAVVLVKDGKVYILVVGEGEKAKKALEEA TKLAKELGFEVEARTIKPEELEEALEEGFRKL	L24
118	SEKARQAYEAARKNFKGGGTVVVLGGEVHAFGGGEEIKAQAEAIREESGG GGSVVVSDNPELALALVEKYTGDPEAVAKAVEELG	L511
119	MVVLTDDGADAAVELVANYTGDPAAVKAALKELTGGGITVVVVFAPDPEL AEKVYKVAEKEAGKDSAEGMKEAAKKIKEETGATDVIVL	L850
120	HMEEIQKAIEELLKTISPEEAAIEIVQRFGAAVVVVTRDLVVVVCSDDHELST RVWEAAQKALETGGERQGKHIAEYIRREFEGEVEVILL	L1090
121	MVVVVGGDTEEGEKFAKIMEEELGGVTTPEERAEKAYEIAVKHYHNAMVL VGVKDPEAVKKAWEELKGDNAKAAEELVEKFGIDVAVV	L186
122	MEEAEKAYKAAKEALEKFGKNKGKKIADYIKKNGGDKFKVIQLEPEIVVVH KGDEKEIKERVEELLKTLSPEEVGLKLAEEFGAVVVVV	L1195
123	AADKVYEIAEKNLGSGEENAKRIAKLVEETLKGEVTVHMFGDVAVFVVGD PEKAAEVRARLEAVEAELKAKGITDPVERRLAFLEAVGAAVVYG	L363
124	MEEEIQKAIEELLKTKSKEEAAIEIVQRFGTAVVVVSSELVVIHGGENHELST RVWEAAQKALETSGERDGKHIAEYIRRYTGGRVEVILI	L76
125	MEEEMKVYEAAREALEKAGENRGRHIAAYIREKAGDKFDVVLVGEDIVVA GKGLDKEAIKKEAEELRKIGLEEAAIKLAEKFGAIVVVV	L643
126	MVVVSTDPEAALTLVEKAGGGDMAEVDKLVAELTKGEPVVVVSGDEVRV FHPTPEAVAVAEAIRKAVGGGNEWEKAEAAWKAAEEAF	L380

Example 3: pH responsive EEP binders that bind and release EEPs

[00304] Fusing EEPs directly to Cas9 may interfere with release of Cas9 from endosomes. In place of fusing EEPs directly to Cas9, pH-responsive EEP binders (e.g., pH-responsive domains) were explored herein that capture EEPs at physiological pH (7.4) and release the EEP as soon as they enter the acidic compartment. Binding scaffolds were first designed with an interface for the peptides (melittin and ppTG21). A histidine network was then introduced with the scaffold to enable the protein to disassemble at acidic pH, thereby releasing the EEP from the pH-responsive EEP binder. pH-responsive EEP binders were designed, ordered, expressed and purified. pH-responsive EEP binders were complexed with their EEP and tested for hemolysis, demonstrating no release at pH 7.4 and incremental release with lower pH (see for example **Table 6** for list of EEP binders tested). These are now under investigation with RNPs on cells.

Table 6. Amino acid sequences of exemplary pH-responsive domains

SEQ ID NO.	Peptide name	Amino acid sequence
336	pHisNet2_ch7	EEEAVEALLRALAALHEALAQRAAADPELARLHREAAE TAEEGVGASREERHELMERHNELARESAKYAERAGNTE HAAIARELIEAAEALMAALK
337	pHisNet2_ch10	EEEAVEALLRALAALHEALARRAAADPELARLHREAAE TAERGVGASREERHELMERHNQLARQSAKYAEEAGNT EHAAIARELIEAAEALMAALK

338	pHisNet2_R22E	EEAAVEALLRALAALHAALAAEAAADPALATLHREAA ATAAAGVGASREERHALMERHNALARASAKYAEAAG NTEHAAIARELIEAAEALMAALK
339	pHisNet2_R22E	EEAAVEALLRALAALHAALAARAAADPALATLHREAA ATAAAGVGASREERHALMERHNALARASAKYAEAAG DTEHAAIARELIEAAEALMAALK
340	pHisNet2_R22E_N74D	EEAAVEALLRALAALHAALAAEAAADPALATLHREAA ATAAAGVGASREERHALMERHNALARASAKYAEAAG DTEHAAIARELIEAAEALMAALK
341	pHisNet2_N59D	EEAAVEALLRALAALHAALAARAAADPALATLHREAA ATAAAGVGASREERHALMERHDALARASAKYAEAAG NTEHAAIARELIEAAEALMAALK
342	pHisNet2_N59D_R63Q	EEAAVEALLRALAALHAALAARAAADPALATLHREAA ATAAAGVGASREERHALMERHDALAQASAKYAEAAG NTEHAAIARELIEAAEALMAALK
343	pHisNet2_N59D_R63Q_E89Q	EEAAVEALLRALAALHAALAARAAADPALATLHREAA ATAAAGVGASREERHALMERHDALAQASAKYAEAAG NTEHAAIARELIEAAQALMAALK
344	pHisNet2_H16N	EEAAVEALLRALAALNAALAARAAADPALATLHREAA ATAAAGVGASREERHALMERHNALARASAKYAEAAG NTEHAAIARELIEAAEALMAALK
345	pHisNet2_H16N_L12I	EEAAVEALLRAIAALNAALAARAAADPALATLHREAAA TAAAGVGASREERHALMERHNALARASAKYAEAAGNT EHAAIARELIEAAEALMAALK
346	pHisNet4	AAAAVDAFFAASAAYHALAAETLPLSPEARALHARARD LAVAAQGASAAEALALLKELLAVHRESAAAHRAAGNE ETAAQEEETAEAVAALLAELA
347	pHisNet5	AAARARELLARMRELAAELLALAEEGVEANRAALLAL HAELAATAAELLREVGASQELVDNHEATAERARAGEL NEEEQEAHAKILAENLEELAR
348	pHisNet2_trunc1	REAALAAAVAALAALHAALAARAAADPALATLHREAA ATAAAGTPELMERHNALARASAKYAEAAGNTEHAAIA RELIEAAEALAEDLA
349	pHisNet2_trunc2	REAALAAARAALAALHAALAARAAADPALATLHREAA ATAAAGTPELMERHNALARASAKYAEAAGNTEHAAIA RELIEAAEALLAELR

Example 4: Endosomal disruption by CRISPR-Cas9 fusions

[00305] Galectin-9 (GAL9) is a known marker of endosomal membrane disruption and can be utilized to quantitate endosomal rupture. Exemplary fusion polypeptides according to some embodiments of the present disclosure (e.g., CRISPR-Cas9-targeting moiety-EEP fusion, or CRISPR-Cas9-targeting moiety-pH responsive domain fusion that is bound with an EEP) are subjected to an mCherry-GAL9-recruitment assay to screen for the ability to induce endosomal rupture in HEK293 cells. Under normal conditions, mCherry-GAL9 is homogenously dispersed in the cytosol. When an endosomal leakage event is detected, mCherry-GAL9 will translocate to the

damaged endosome. The resulting localization pattern of discrete puncta can be visualized and quantitated.

[00306] Briefly, for the GAL9 puncta assay, HEK293 cells are transduced with lentivirus containing a mCherry-GAL9 fusion protein under a CMV promoter and a puromycin resistance gene. Pooled stable cell lines are established by Puromycin selection for 48h and cells are maintained in Puromycin media thereafter.

[00307] HEK293 mCherry-Gal9 cells are incubated with the exemplary fusion proteins in complete media (DMEM supplemented with 10% fetal bovine serum and 1% Pens/Strep) for up to 4h before adding Draq5 DNA stain to stain nuclei of live cells. GAL9 is recruited to sites of endosomal damage resulting in puncta formation visible by microscopy. The cells are imaged using an ImageXpress high-content confocal imager (Molecular Devices) using a 20x objective. Using InCarta image analysis software, Gal9 puncta per cell are calculated, demonstrating that exemplary fusion proteins of the present disclosure can induce endosomal rupture.

Example 5: GFP editing by CRISPR-Cas9 fusions

[00308] Green fluorescent protein (GFP) is extensively utilized as a reporter and is easily detectable in living cells. Nonetheless, its high stability (e.g., long half-life) results in the accumulation of GFP in cells, which limits its usefulness in assays that necessitate swift turnover responses (e.g., in signal pathways) and in knockdown or knockout detection. Unstable GFP (uGFP) has a half-life of ~2h and is degraded quickly upon introducing frameshifts to the coding sequence (e.g., by CRISPR-Cas9). Given it's instability, uGFP is suitable for observing time course induction and dose-response kinetics, as well as rapid detection of knockdown or knockout assays. Thus, GFP fluorescence of cells 48h-72h after gene editing (e.g., with CRISPR-Cas9) can be used as a gene editing outcome proxy.

[00309] Briefly, HEK293 cells are transduced with lentivirus containing an unstable GFP protein (uGFP) under a CMV promoter and a puromycin resistance gene. Stable cell lines are established by Puromycin selection for 48h and cells are maintained in Puromycin media thereafter.

For self-delivery experiments, HEK293-uGFP cells are seeded at 1e4 cells/well in Poly-D-lysine coated 96-well plates. The day after, exemplary fusion polypeptides according to some embodiments of thre present disclosure (e.g., CRISPR-Cas9-targeting moiety-EEP fusion, or CRISPR-Cas9-targeting moiety-pH responsive domain fusion that is bound with an EEP) are added to the cells into complete media for 4h at 37C and 5% CO2. After 4h, media is removed and replaced with pre-warmed complete growth media for 48h-72h. The cells are then imaged using an ImageXpress high-

content confocal imager (Molecular Devices) using a 4x objective. Using InCarta image analysis software, % GFP negative cells are calculated.

Example 6: Biodistribution of CRISPR-Cas9 fusions

[00310] The biodistribution of exemplary fusion polypeptides according to some embodiments of the present disclosure (e.g., CRISPR-Cas9-targeting moiety-EEP fusion, or CRISPR-Cas9-targeting moiety-pH responsive domain fusion that is bound with an EEP) is assessed in a mouse model.

[00311] Briefly, biodistribution of the exemplary fusion polypeptides is evaluated by administering doses of 5 mg/kg or 10 mg/kg via intraperitoneal injection in mice. At predefined timepoints (e.g., 15 minutes, 1 hour, and 3 hours post-administration), tissues, including liver, kidney, spleen, lung, brain, heart, gastrocnemius muscle and blood, are collected to assess compound distribution and pharmacokinetic profiles. Quantification of fusion peptide concentrations in tissues are performed using a standard LC/MS assay, separating analytes on a reverse phase column after precipitation in serum, followed by identifying each parent ion and quantifying the daughter ion for each analyte.

Example 7: *In vivo* editing by CRISPR-Cas9 fusions

[00312] The *in vivo* editing by exemplary fusion polypeptides according to some embodiments of the present disclosure (e.g., CRISPR-Cas9-targeting moiety-EEP fusion, or CRISPR-Cas9-targeting moiety-pH responsive domain fusion that is bound with an EEP) is assessed in a mouse model.

[00313] Briefly, mice are dosed intravenously with 10 mg/kg of the exemplary fusion polypeptides. Tissues are collected 7 days post-injection and analyzed using next-generation sequencing (NGS) to assess editing efficiency of the fusion peptides on an endogenous gene.

Example 8: Pharmacokinetic of CRISPR-Cas9 fusions

[00314] The pharmacokinetic activity of exemplary fusion polypeptides according to some embodiments of the present disclosure (e.g., CRISPR-Cas9-targeting moiety-EEP fusion, or CRISPR-Cas9-targeting moiety-pH responsive domain fusion that is bound with an EEP) is assessed in an *in vitro* model.

[00315] Briefly, the exemplary fusion polypeptides are prepared and utilized to evaluate kinetic solubility, plasma protein binding (PPB), logD, stability in primary mouse hepatocytes, and permeability in MDR1-MDCK cell lines.

[00316] Plasma protein binding is measured using an equilibrium dialysis method. Briefly, dialysis membranes are prepared, and the exemplary fusion polypeptides are spiked into mouse plasma to achieve a final concentration of 1.0 μM. The samples are incubated with the membranes and then

collected from both compartments after a specified incubation period. The samples are analyzed using LC-MS/MS.

[00317] To assess stability in primary mouse hepatocytes, briefly, cryopreserved hepatocytes are thawed, diluted to a concentration of 1 million cells/mL in Williams E Medium, and assessed for viability. Incubations are performed in duplicate within 96-well plates containing the exemplary fusion polypeptides **or a positive control at 1 μM.** The samples are analyzed using LC-MS/MS.

[00318] Permeability assays in MDR1-MDCK cell lines are conducted using a 24-well plate format. Briefly, cells are seeded four days prior to the assay, and their integrity is confirmed via transepithelial electrical resistance (TEER) measurements. Donor and receiver solutions containing varying concentrations of DMSO and Lucifer Yellow (5 μM) are prepared according to the transport direction (A-to-B or B-to-A). After washing the cell monolayers and measuring TEER values again, the plates are incubated with the exemplary fusion polypeptides at 37°C for 90 minutes. Post-incubation, samples from both apical and basolateral chambers are collected for analysis of Lucifer Yellow concentrations using a fluorometer, as well as for LC-MS/MS preparation.

[00319] 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. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. 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.

CLAIMS

What is claimed is:

1. A composition comprising a fusion polypeptide comprising:
   (a) an effector polypeptide; and
   (b) a targeting moiety that specifically binds to an extracellular domain of a receptor expressed by a target cell, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody.
2. The composition of claim 1, wherein the fusion polypeptide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.
3. The composition of claim 1 or 2, wherein the fusion polypeptide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.
4. A composition comprising a fusion polypeptide comprising:
   (a) an effector polypeptide; and
   (b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 31-151** or **350-143854**.
5. The composition of claim 4, wherein the fusion polypeptide further comprises a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.
6. The composition of claim 4 or 5, wherein the fusion polypeptide further comprises an endosomal escape peptide, wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.
7. A composition comprising a fusion polypeptide comprising:
   (a) an effector polypeptide; and
   (b) a pH responsive polypeptide domain, wherein the pH responsive polypeptide domain is capable of binding to an endosomal escape peptide at pH 7.4, and wherein when the fusion

polypeptide is present at a pH lower than 7.0, the pH responsive polypeptide domain contains one or more amino acids that are protonated.

8. A composition comprising a fusion polypeptide comprising:
(a) an effector polypeptide; and
(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 336-349**.
9. The composition of claim 7 or 8, wherein the fusion polypeptide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell.
10. The composition of claim 9, wherein the targeting moiety is not an antibody or an antigen binding domain of an antibody.
11. A composition comprising a fusion polypeptide comprising:
(a) an effector polypeptide; and
(b) an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 152-335**.
12. A composition comprising a protein complex, the protein complex comprising:
(a) the fusion polypeptide of the composition of claim 7 or 8, and
(b) an endosomal escape peptide, wherein the endosomal escape peptide is bound to the fusion polypeptide, and wherein the endosomal escape peptide is capable of disrupting a membrane of a membrane-bound organelle.
13. The composition of claim 11 or 12, wherein the fusion polypeptide further comprises a targeting moiety that binds specifically to an extracellular domain of a receptor expressed by a target cell.
14. The composition of claim 13, wherein the endosomal escape peptide is bound to a pH responsive polypeptide domain.
15. The composition of any one of claims 1-3, 9-10 and 13-14, wherein the targeting moiety binds to an extracellular domain of transferrin receptor protein 1 (TfR1), IGF1R, IGF2R, Dnmt3a, or Nephrin.
16. The composition of any one of claims 1-3, 9-10 and 13-15, wherein the targeting moiety comprises a polypeptide sequence with less than 150 amino acids.
17. The composition of any one of claims 1-3, 9-10 and 13-16, wherein the targeting moiety comprises a polypeptide sequence with less than 100 amino acids.

18. The composition of claim 16 or 17, wherein the targeting moiety comprises a polypeptide sequence with more than 50 amino acids.
19. The composition of claim 18, wherein the targeting moiety comprises a polypeptide sequence with more than 60 amino acids.
20. The composition of claim 18, wherein the targeting moiety comprises a polypeptide sequence with more than 70 amino acids.
21. The composition of claim 18, wherein the targeting moiety comprises a polypeptide sequence with 80 amino acids or more.
22. The composition of any one of claims 1-3, 9-10 and 13-21, wherein the targeting moiety comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence any one of **SEQ ID NOs: 31-151** or **350-143854**.
23. The composition of any one of claims 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated at a pH of 6.8 or less, and wherein the one or more amino acids that are protonated at a pH of 6.8 or less are not protonated at a pH of 7.4.
24. The composition of any one of claims 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 6.6 or less, and wherein the one or more amino acids that are protonated at a pH of 6.6 or less are not protonated at a pH of 7.4.
25. The composition of any one of claims 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 6 or less, and wherein the one or more amino acids that are protonated at a pH of 6 or less are not protonated at a pH of 7.4.
26. The composition of any one of claims 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5.5 or less, and wherein the one or more amino acids that are protonated at a pH of 5.5 or less are not protonated at a pH of 7.4.
27. The composition of any one of claims 2, 3, 5-7, and 14-22, wherein the pH responsive polypeptide domain contains one or more amino acids are protonated at a pH of 5 or less, and wherein the one or more amino acids that are protonated at a pH of 5 or less are not protonated at a pH of 7.4.

28. The composition of any one of claims 2, 3, 5-7, and 14-27, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polypeptide is present within an endosome.
29. The composition of any one of claims 2, 3, 5-7, and 14-28, wherein the pH responsive polypeptide domain contains one or more amino acids that are protonated when the fusion polypeptide is present within a lysosome.
30. The composition of any one of claims 2, 3, 5-7, and 14-29, wherein the pH responsive polypeptide domain comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 336-349**.
31. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
32. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
33. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
34. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
35. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
36. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 6.8 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold

less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.

37. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 6.6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
38. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 6 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
39. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 5.5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
40. The composition of any one of claims 2, 3, 5-7, and 14-30, wherein at a pH of 5 or less the endosomal escape peptide has an affinity to the pH responsive polypeptide domain that is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 500 or 1000 fold less than an affinity of the endosomal escape peptide to the pH responsive polypeptide domain at a pH of 7.4.
41. The composition of any one of claims 3, 6, and 12-40, wherein the endosomal escape peptide is released from the protein complex when the protein complex is present within an endosome.
42. The composition of any one of claims 3, 6, and 12-41, wherein the endosomal escape peptide is released from the protein complex when the protein complex is present within a lysosome.
43. The composition of any one of claims 3, 6, and 12-42, wherein the endosomal escape peptide comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of **SEQ ID NOs: 152-335**.
44. The composition of any one of claims 2, 3, 5-7, and 14-43, wherein the pH responsive polypeptide domain is capable of binding to two or more endosomal escape peptides at pH 7.4.

45. The composition of any one of claims 12-44, wherein the protein complex comprises two or more copies of the endosomal escape peptide.
46. The composition of any one of claims 1-45, wherein the fusion polypeptide further comprises a linker linking the effector polypeptide and one or more other peptide sequences of the fusion polypeptide.
47. The composition of claim 46, wherein the linker is a proteolytically cleavable linker.
48. The composition of any one of claims 1-47, wherein the fusion polypeptide further comprises one or more nuclear localization sequences (NLSs).
49. The composition of claim 48, wherein the one or more NLSs comprise the amino acid sequence K(K/R)X(K/R), and where X is any amino acid.
50. The composition of claim 48, wherein the one or more NLSs comprise the amino acid sequence PKKKRKV.
51. The composition of any one of claims 1-50, wherein the one or more NLSs are at the N-terminus of the effector polypeptide.
52. The composition of any one of claims 1-51, wherein the effector polypeptide is a CRISPR-Cas effector polypeptide.
53. The composition of any one of claims 1-51, wherein the CRISPR-Cas effector polypeptide is a Type II CRISPR-Cas effector polypeptide, a Type V CRISPR-Cas effector polypeptide, or a Type VI CRISPR-Cas effector polypeptide.
54. The composition of any one of claims 1-51, wherein the CRISPR-Cas effector polypeptide is a SpyCas9 or a variant thereof.
55. The composition of any one of claims 1-51, wherein the CRISPR-Cas effector polypeptide is GeoCas9 or a variant thereof.
56. The composition of any one of claims 1-55, wherein the CRISPR-Cas effector polypeptide is catalytically active.
57. The composition of any one of claims 1-56, wherein the CRISPR-Cas effector polypeptide exhibits reduced catalytic activity compared to a wild-type CRISPR-Cas effector polypeptide.
58. The composition of any one of claims 1-55, wherein the CRISPR-Cas effector polypeptide is catalytically inactive.
59. The composition of any one of claims 1-58, wherein the fusion polypeptide further comprises at least one additional heterologous polypeptide.
60. The composition of claim 59, wherein the at least one additional heterologous polypeptide is a deaminase, a base editor, a reverse transcriptase, a transcription modulator, or an epigenetic modulator.

61. A composition comprising:

(a) a fusion polypeptide of the composition of any one of claims 1-60, or a nucleic acid comprising a nucleotide sequence encoding the fusion polypeptide and/or endosomal escape polypeptide; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

62. The composition of claim 61, wherein the guide nucleic acid is a single-molecule guide nucleic acid.

63. The composition of claim 61, further comprising a donor nucleic acid.

64. A nucleic acid comprising a nucleotide sequence encoding the fusion polypeptide of the composition of any one of claims 1-60.

65. The nucleic acid of claim 64, wherein the nucleotide sequence is operably linked to a transcriptional control element, optionally wherein the transcriptional control element is a promoter.

66. A recombinant expression vector comprising the nucleic acid of claim 65.

67. A cell comprising the composition of any one of claims 1-63 or the nucleic acid of claim 64 or 65, or the recombinant expression vector of claim 66.

68. The cell of claim 67, wherein the cell is a eukaryotic cell.

69. The cell of claim 67 or 68, wherein the cell is *in vitro*.

70. The cell of claim 67 or 68, wherein the cell is *in vivo*.

71. A method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with:

(a) the composition of any one of claims 1-63; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

72. A method of modifying a target nucleic acid in a eukaryotic cell, the method comprising introducing into the eukaryotic cell:

(a) the composition of any one of claims 1-63; and

(b) a guide nucleic acid, or a nucleic acid comprising a nucleotide sequence encoding the guide nucleic acid.

73. The method of claim 72, further comprising introducing into the cell a donor nucleic acid.

74. The method of claim 72 or 73, wherein the cell is *in vitro*.

75. The method of claim 72 or 73, wherein the cell is *in vivo*.

76. The method of any one of claims 72-75, wherein the cell is a mammalian cell, an insect cell, an avian cell, a reptile cell, an amphibian cell, an arachnid cell, a protozoan cell, or a plant cell.
77. The method of any one of claims 72-76, wherein the target nucleic acid is selected from: double stranded DNA, single stranded DNA, RNA, genomic DNA, and extrachromosomal DNA.
78. The method of any one of claims 72-77, wherein the modifying comprises genome editing.