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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 June 2025 (26.06.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/137461 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/113* (2010.01)
A61K 38/46 (2006.01) *C12N 15/52* (2006.01)
A61K 9/127 (2025.01) *C12N 15/86* (2006.01)
C07K 14/195 (2006.01) *C12N 5/10* (2006.01)
C12N 15/10 (2006.01) *A61K 35/12* (2015.01)

(21) International Application Number:
PCT/US2024/061295

(22) International Filing Date:
20 December 2024 (20.12.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/612,882 20 December 2023 (20.12.2023) US

(71) Applicant: **IONIS PHARMACEUTICALS, INC.** [US/US]; 2855 Gazelle Ct., Carlsbad, California 92010 (US).

(72) Inventors: **ZHAO, Jing**; c/o Ionis Pharmaceuticals, Inc., 2855 Gazelle Ct., Carlsbad, California 92010 (US). **RIGO, Frank**; c/o Ionis Pharmaceuticals, Inc., 2855 Gazelle Ct., Carlsbad, California 92010 (US). **ECKER, David J.**; c/o Ionis Pharmaceuticals, Inc., 2855 Gazelle Ct., Carlsbad, California 92010 (US). **MICHAEL, Todd**; 2410 Copper Way, Carlsbad, California 92009 (US). **LI, Zheng**; c/o Ionis Pharmaceuticals, Inc., 2855 Gazelle Ct., Carlsbad, California 92010 (US).

(74) Agent: **NEMIROW, Joy Lynn**; Sheppard Mullin Richter & Hampton LLP, 12275 El Camino Real, Suite 100, San Diego, California 92130 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: NUCLEIC ACID BINDING AGENTS AND USES THEREOF

(57) Abstract: Provided are guided nucleic acid binding agents, systems comprising guided nucleic acid binding agents, guides, and methods of use thereof. Such guided nucleic acid binding agents and systems are useful for gene editing.

WO 2025/137461 A1

NUCLEIC ACID BINDING AGENTS AND USES THEREOF

Sequence Listing

The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled EDIT0003WOSEQ.xml created on December 19, 2024, which is 1,883 KB in size. The information in the electronic format of the sequence listing is incorporated herein by reference in its entirety.

Field

The present embodiments provide nucleic acid binding agents, systems comprising nucleic acid binding agents, and methods of use thereof.

Background

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. These CRISPR-RNA guided nucleic acid cleavage systems serve to protect such microorganisms against non-self nucleic acids, such as infectious viruses and plasmids. CRISPR-associated (Cas) proteins are highly diverse, containing a wide variety of nucleic acid-binding domains. Cas nucleic acid-binding domains can be harnessed for a variety of applications, including, but not limited to, gene editing. Specific applications include, but are not limited to, gene silencing, gene activation, homologous gene repair, gene visualization, base editing, prime editing and epigenetic modulation of gene expression.

CRISPR-Cas systems can be divided in two classes, with Class 1 systems utilizing a complex of multiple Cas proteins (such as type I, III, and IV CRISPR-Cas systems) and Class 2 systems utilizing a single Cas protein (such as Type II, V, and VI CRISPR-Cas systems). Type II CRISPR-Cas-based systems have been used for genome editing, and require a Cas protein or variant thereof guided by a customizable guide for programmable DNA targeting.

Few Class 2 CRISPR/Cas systems have been widely used. Of these, Type V are notable by utilizing a single unified RuvC-like endonuclease (RuvC) domain that recognizes 5' PAM sequences that are different from the 3' PAM sequences recognized by the more widely used Cas9 (a Type II system), and form a staggered cleavage in the target nucleic acid with 5, 7, or 10 nucleotide 5′ overhangs (Yang et al., PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. Cell 167:1814 (2016)). However, wild-type Type V Cas nuclease and guide sequences typically have low editing efficiency. In addition, challenges remain with existing systems, such as efficient targeting, reducing immunogenicity,

reducing accelerated blood clearance (ABC) and reducing dose-limiting toxicity such as acute phase response (APR) and complement activation-related pseudoallergy (CARPA). Accordingly, there is an ongoing need for additional Class 2, Type V CRISPR/Cas systems that offer improvements over prior art systems in a variety of therapeutic, diagnostic, and research applications.

Summary

Certain embodiments provided herein are directed to guided nucleic acid binding agents, editing systems comprising guided nucleic acid binding agents, and methods of use thereof.

In certain embodiments, provided herein are polypeptides. In certain embodiments, polypeptides are guided nucleic acid binding agents. In certain embodiments, polypeptides provided herein are Cas proteins. In certain embodiments, the guided nucleic acid binding agent is a polypeptide comprising a RuvC domain. In certain embodiments, the guided nucleic acid binding agent is a polypeptide containing exactly one RuvC domain. In certain embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain embodiments, the guided nucleic acid binding agent is a Cas protein. In certain such embodiments, a Cas protein is a Cas enzyme and is capable of inducing cleavage of one strand of a double-stranded DNA target (a nick). In certain embodiments, the guided nucleic acid binding agent is a Class 2, Type V Cas endonuclease. In certain embodiments, the guided nucleic acid binding agent comprises a Class 2, Type V Cas endonuclease. In certain embodiments, the guided nucleic acid binding agent contains exactly one nuclease active site. In certain embodiments, the guided nucleic acid binding agent lacks nuclease activity and comprises a "dead Cas" or "dCas". In certain embodiments, the dead Cas is a modified Cas polypeptide with an inactivating substitution in the nuclease active site. In certain embodiments, polypeptides comprise Cas proteins. Certain Cas proteins may be fused with one or more other proteins or protein domains, also known as 'heterologous domains', to create Cas fusion proteins.

In certain embodiments, provided herein are nucleic acids that encode the polypeptides described above. In certain embodiments, such nucleic acids are DNA (e.g., template DNA) or RNA (e.g., mRNA).

In certain embodiments, provided herein are guides. In certain embodiments, a guide consists of a single oligonucleotide (a single guide). Guides comprise a target-recognition region and a protein-recognition region. In certain embodiments, the single guide is an oligonucleotide that comprises a target-recognition region and a protein-recognition region. In certain embodiments, the target-recognition region is a region of an oligonucleotide that has a sequence complementary to an equal length portion of a target sequence of a target nucleic acid. In certain embodiments, the target nucleic acid is DNA.

In certain embodiments, an editing system comprises a guided nucleic acid binding agent and a guide. In certain such embodiments, the guided nucleic acid binding agent is a Cas protein. In certain embodiments, the guide is a single guide. In certain embodiments, an editing system comprises a Cas protein and a single guide. In certain embodiments, the editing system comprises a ribonucleoprotein (RNP). In certain embodiments, the RNP comprises a guided nucleic acid binding agent and a guide.

In certain embodiments, there is provided an expression system comprising a nucleic acid such as an exogenous mRNA encoding a guided nucleic acid binding agent, and a guide. In certain such embodiments, the guided nucleic acid binding agent is a Cas protein. In certain embodiments, the guide is a single guide.

In certain embodiments, there is provided a viral vector comprising a nucleic acid encoding a guided nucleic acid binding agent, and a guide.

In certain embodiments, there is provided a lipid nanoparticle (LNP) comprising an editing system. In certain embodiments, the editing system comprises a guided nucleic acid binding agent and a guide. In certain such embodiments, the guided nucleic acid binding agent is a Cas protein. In certain embodiments, the guide is a single guide. In certain embodiments, an editing system comprises a Cas protein and a single guide. In certain embodiments, the editing system comprises a ribonucleoprotein (RNP). In certain embodiments, the RNP comprises a guided nucleic acid binding agent and a guide. In certain embodiments, the RNP comprises a Cas protein and a guide.

In certain embodiments, there is provided a lipid nanoparticle (LNP) comprising an expression system. In certain embodiments, the expression system comprises a nucleic acid and a guide. In certain embodiments, the nucleic acid is an exogenous mRNA encoding a guided nucleic acid binding agent. In certain such embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain embodiments, the guide is a single guide.

In certain embodiments, there is provided a composition comprising an LNP, wherein the LNP comprises an editing system. In certain embodiments, the editing system comprises a guided nucleic acid binding agent and a guide. In certain such embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain embodiments, the guide is a single guide. In certain embodiments, an editing system comprises a Cas protein and a single guide. In certain embodiments, the editing system comprises a ribonucleoprotein (RNP). In certain embodiments, the RNP comprises a guided nucleic acid binding agent and a guide. In certain embodiments, the RNP comprises a Cas protein and a guide.

In certain embodiments, there is provided a composition comprising an editing system. In certain embodiments, the editing system comprises a guided nucleic acid binding agent and a guide. In certain such embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain embodiments, the guide is a single guide. In certain embodiments, an editing system comprises a Cas protein and a single guide. In certain embodiments, the editing system comprises a ribonucleoprotein (RNP). In certain embodiments, the RNP comprises a guided nucleic acid binding agent and a guide. In certain embodiments, the RNP comprises a Cas protein and a guide.

In certain embodiments, there is provided a composition comprising a ribonucleoprotein (RNP). In certain embodiments, the RNP comprises a guided nucleic acid binding agent and a guide. In certain embodiments, the RNP comprises a Cas protein and a guide.

In certain embodiments, there is provided a composition comprising an expression system. In certain embodiments, the composition comprises a viral vector comprising the expression system. In certain

embodiments, the expression system comprises a nucleic acid and a guide. In certain embodiments, the viral vector comprises a nucleic acid encoding a guided nucleic acid binding agent, and/or a guide. In certain embodiments, the nucleic acid is an exogenous mRNA encoding a guided nucleic acid binding agent. In certain such embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain embodiments, the guide is a single guide.

In certain embodiments, there is provided a method of editing a target nucleic acid, a method of creating a discontinuity in a target nucleic acid, a method of creating a double stranded break or a nick in a target DNA, a method of gene silencing, a method of gene activation, a method of homologous gene repair, a method of gene visualization, a method of base editing, a method of prime editing or a method of epigenetic modulation of gene expression. In such embodiments, the method comprises administering an LNP, a composition, an editing system, an expression system, a viral vector, or a guided nucleic acid binding agent and a guide disclosed herein.

In certain embodiments, there is provided a cell comprising an edited target nucleic acid, a discontinuity in a target nucleic acid, a double stranded break or a nick in a target DNA, a silenced gene, an activated gene, a homologously repaired gene, a gene that can be visualized, or a gene the expression of which has been epigenetically modified, produced by the methods disclosed herein.

In certain embodiments, there is provided a method of treating a disease or disorder in a subject, a method of autologous cell therapy or a method of allogeneic cell therapy. In such embodiments, the method comprises administering an LNP, a composition, an editing system, an expression system, or a viral vector as disclosed herein. In certain embodiments, the LNP, composition, editing system, expression system, or viral vector comprises a nucleic acid encoding a guided nucleic acid binding agent or a guided nucleic acid binding agent, and a nucleic acid encoding a guide or a guide.

In certain embodiments, there is provided use of an LNP, a composition, an editing system, an expression system, or a viral vector disclosed herein in the manufacture of a medicament for treating a disease or disorder in a subject, including for autologous cell therapy or for allogeneic cell therapy. In certain embodiments, there is provided use of a guided nucleic acid binding agent and a guide disclosed herein in the manufacture of a medicament for treating a disease or disorder in a subject, including for autologous cell therapy or for allogeneic cell therapy. In certain such embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain such embodiments, the guide is a single guide.

In certain embodiments, there is provided a method of diagnosing a disease or disorder in a subject, or a method of assessing responsiveness to a treatment of a disease or disorder in a subject, wherein the method comprises administering an LNP, a composition, an editing system, an expression system, or a viral vector disclosed herein. In certain embodiments, there is provided a method of diagnosing a disease or disorder in a subject, or a method of assessing responsiveness to a treatment of a disease or disorder in a subject, wherein the method comprises administering a guided nucleic acid binding agent and a guide as

disclosed herein. In certain such embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain such embodiments, the guide is a single guide. .

Further disclosed herein are kits and devices comprising an LNP, a composition, an editing system, an expression system, or a viral vector as herein disclosed, and optionally instructions for use. Further disclosed herein are kits and devices comprising a guided nucleic acid binding agent and a guide as herein disclosed, and optionally instructions for use. In certain such embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain such embodiments, the guide is a single guide.

Brief Description of the Drawings

Figure 1 shows gels showing cleavage activity of Cas proteins in HEK293T nuclear extracts using guides having a general target-recognition region and a Cas-specific protein-recognition region in an embodiment according to Example 2.

Figure 2a and Figure 2b show gels showing cleavage activity of Cas proteins in HEK293T nuclear extracts using guides having a general target-recognition region and a protein-recognition region specific to AsCas12a in an embodiment according to Example 2.

Figure 3 shows a guide interacting with its target, labeled with its protein-recognition region, direct repeat region, target-recognition region, and 3′-extension.

Detailed Description

It is to be understood that both the foregoing background and summary, and the following detailed description are exemplary and explanatory only and are not restrictive of the embodiments, as claimed. Herein, the use of the singular includes the plural unless specifically stated otherwise. As used herein, the use of "or" means "and/or" unless stated otherwise. Furthermore, the use of the term "including" as well as other forms, such as "includes" and "included", is not limiting.

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

All documents, or portions of documents, cited in this application, including, but not limited to, patents, patent applications, articles, books, treatises, and GenBank and NCBI reference sequence records, including any drawings and appendices, are hereby expressly incorporated by reference for the portions of the document discussed herein, as well as in their entirety for all purposes to the same extent as if each was specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgement or any form of suggestion that prior art forms part of the common general knowledge of the person skilled in the art.

As used herein, numerical ranges are provided for certain quantities. It is to be understood

that these ranges comprise all subranges therein. Thus, the range "from 50 to 80" includes all possible ranges therein (e.g., 51-79, 52-78, 53-77, 54-76, 55-75, 60-70, etc.). Furthermore, all values within a given range may be an endpoint for the range encompassed thereby (e.g., the range 50-80 includes the ranges with endpoints such as 55-80, 50-75, etc.). It is understood that throughout the specification, the first letter in a peptide sequence is the first amino acid of the peptide at the N-terminus and the last letter in a peptide sequence is the last amino acid of the peptide at the C-terminus unless indicated otherwise. Similarly, the first nucleoside in a nucleotide sequence represents the 5'-end of the nucleotide, and the last letter in the nucleotide sequence represents the 3' end, unless indicated otherwise.

Unless specific definitions are provided, the nomenclature used in connection with, and the procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Where permitted, all patents, applications, published applications and other publications and other data referred to throughout in the disclosure are incorporated by reference herein in their entirety.

Unless otherwise indicated, the following terms have the following meanings:

As used herein, "2'-deoxynucleoside" means a nucleoside comprising a 2'-H(H) deoxyfuranosyl sugar moiety. In certain embodiments, a 2'-deoxynucleoside is a 2'-β-D-deoxynucleoside and comprises a 2'-β-D-deoxyribosyl sugar moiety, which has the β-D ribosyl configuration as found in naturally occurring deoxyribonucleic acids (DNA). In certain embodiments, a 2'-deoxynucleoside may comprise a modified nucleobase or may comprise an RNA nucleobase (uracil).

As used herein, "2'-MOE" means a 2'-$OCH_2CH_2OCH_3$ group in place of the 2'-OH group of a furanosyl sugar moiety. A "2'-MOE sugar moiety" means a sugar moiety with a 2'-$OCH_2CH_2OCH_3$ group in place of the 2'-OH group of a furanosyl sugar moiety. Unless otherwise indicated, a 2'-MOE sugar moiety is in the β-D-ribosyl configuration. "MOE" means O-methoxyethyl.

As used herein, "2'-MOE nucleoside" means a nucleoside comprising a 2'-MOE sugar moiety.

As used herein, "2'-OMe" means a 2'-OCH_3 group in place of the 2'-OH group of a furanosyl sugar moiety. A "2'-O-methyl sugar moiety" or "2'-OMe sugar moiety" means a sugar moiety with a 2'-OCH_3 group in place of the 2'-OH group of a furanosyl sugar moiety. Unless otherwise indicated, a 2'-OMe sugar moiety is in the β-D-ribosyl configuration.

As used herein, "2'-OMe nucleoside" means a nucleoside comprising a 2'-OMe sugar moiety.

As used herein, "2'-F" means a 2'-F group in place of the 2'-OH group of a furanosyl sugar moiety. A "2'-fluoro sugar moiety" or "2'-F sugar moiety" means a sugar moiety with a 2'-F group in place of the 2'-OH group of a furanosyl sugar moiety. Unless otherwise indicated, a 2'-F sugar moiety is in the β-D-ribosyl configuration.

As used herein, "2'-F nucleoside" means a nucleoside comprising a 2'-F sugar moiety.

As used herein, "2'-NMA" means a $-O-CH_2-C(=O)-NH-CH_3$ group in place of the 2'-OH group of a ribosyl sugar moiety. A "2'-NMA sugar moiety" is a sugar moiety with a $2'-O-CH_2-C(=O)-NH-CH_3$ group in place of the 2'-OH group of a ribosyl sugar moiety. Unless otherwise indicated, a 2'-NMA sugar moiety is in the β-D configuration. "NMA" means O-(N-methyl)acetamide.

As used herein, "2'-NMA nucleoside" means a nucleoside comprising a 2'-NMA sugar moiety.

As used herein, "2'-substituted nucleoside" means a nucleoside comprising a 2'-substituted sugar moiety. As used herein, "2'-substituted" in reference to a sugar moiety means a sugar moiety comprising at least one 2'-substituent group other than H or OH.

As used herein, "5-methyl cytosine" means a cytosine modified with a methyl group attached to the 5 position. A 5-methyl cytosine is a modified nucleobase.

As used herein, "active site" means a region of a protein where a substrate molecule binds to undergo enzymatic catalysis of a chemical reaction. An active site comprises one or more catalytic residues. As used herein, "catalytic residue" means an amino acid that is involved in the enzymatic catalysis of a chemical reaction at its active site. A catalytic residue may be directly involved in the catalytic mechanism (e.g., as a nucleophile), may exert an effect on another residue or water molecule which is directly involved in the catalytic mechanism which aids catalysis (e.g., by electrostatic or acid-base action), may stabilize a transition-state intermediate, and/or may exert an effect on a substrate or cofactor which aids catalysis (e.g., by polarizing a bond which is to be broken), including steric and electrostatic effects.

As used herein, "administration" or "administering" refers to routes of introducing a compound or composition provided herein to a subject to perform its intended function. Examples of routes of administration that can be used include, but are not limited to, administration by inhalation, parenteral routes (e.g. subcutaneous injection, intramuscular injection, intravenous infusion, intraarterial infusion, intrathecal injection), topical administration and oral administration.

As used herein, "cargo" refers to an agent that is at least partially encapsulated in a lipid nanoparticle and at least partially released after the lipid nanoparticle is taken up by a cell. Cargo may include, but is not limited to, nucleic acids, oligonucleotides, peptides, proteins, and small molecules. An LNP may contain one or more types of cargo. In certain embodiments, cargo may include a guide, a guided nucleic acid binding agent and/or a ribonucleoprotein.

As used herein, "Cas enzyme" means a guided nucleic acid binding agent that is capable of cleaving one or both strands of double-stranded DNA. In certain embodiments, the target nucleic acid is a double-stranded "target DNA". A "Cas enzyme" may be an enzyme capable of cleaving both strands of a double-

stranded DNA or may be an enzyme capable of cleaving only one strand of a double-stranded DNA (a "Cas nickase"), (e.g., containing a mutation in one but not both of the catalytic residues).

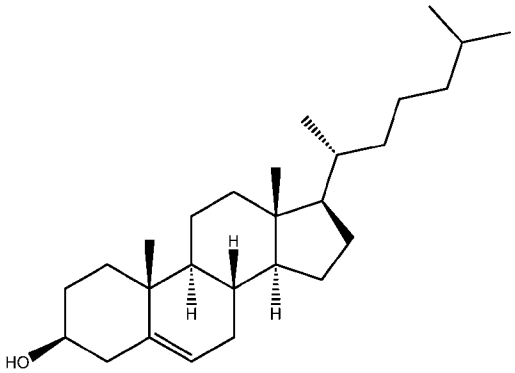
As used herein, a "Cas protein" means a "Cas enzyme" or a "dead Cas protein".

As used herein, a "dead Cas protein" means a guided nucleic acid binding agent that cannot cleave either strand of a double-stranded DNA.

As used herein, a "cationic lipid" is a lipid molecule that carries a nonexchangeable net positive charge.

As used herein, "cEt" or "constrained ethyl" means a bicyclic sugar moiety, wherein the first ring of the bicyclic sugar moiety is a ribosyl sugar moiety, the second ring of the bicyclic sugar is formed via a bridge connecting the 4'-carbon and the 2'-carbon, the bridge has the formula 4'-CH(CH_3)-O-2', and the bridge is in the *S* configuration. A cEt bicyclic sugar moiety is in the β-D configuration.

As used herein, "cholesterol" means:



cholesterol

As used herein, "cleavable moiety" means a bond or group of atoms that is cleaved under physiological conditions, for example, inside a cell, or a subject.

As used herein, "complementary" in reference to an oligonucleotide means that at least 70% of the nucleobases of such oligonucleotide or one or more regions thereof and the nucleobases of another nucleic acid or one or more regions thereof are capable of hydrogen bonding with one another when the nucleobase sequence of the oligonucleotide and the other nucleic acid are aligned in opposing directions. Complementary nucleobases are nucleobase pairs that are capable of forming hydrogen bonds with one another. Complementary nucleobase pairs include adenine (A) and thymine (T); adenine (A) and uracil (U),; cytosine (C) and guanine (G); and 5-methyl cytosine (^{m}C) and guanine (G).

As used herein, the "complementary strand" of a target DNA is the strand of a double-stranded DNA that is complementary to the target-recognition region of a guide. As used herein, the "non-complementary strand" of a target DNA is the strand of a double-stranded DNA that is complementary to the

"complementary strand". The PAM sequence is found within the "non-complementary strand".

As used herein, a "direct repeat region" means the region of a guide for a Class II, Type V Cas enzyme that forms a hairpin to the 5' of the target-recognition region. A "direct repeat region" is part or all of the "protein-recognition region".

As used herein, "domain" means a subset of linked amino acids in a polypeptide. A domain may be comprised of multiple smaller units called subdomains, for example the RuvC domain of *Streptococcus pyogenes* SpyCas9 is comprised of three discontiguous subdomains: RuvC-I, RuvC-II, and RuvC-III.

As used herein, "double-stranded break", "double-strand break", or "DSB" refers to a break in both strands of a double-stranded DNA.

As used herein, "editing system" is a system comprising at least one guided nucleic acid binding agent and at least one guide.

As used herein, "effective amount" means the amount of a formulation according to the invention that, when administered to an animal, is sufficient to effect desired treatment. The "effective amount" will vary depending on the active ingredient, the state, disorder, or condition to be treated and its severity, and the age, weight, physical condition and responsiveness of the animal to be treated.

As used herein, "endonuclease" and "nuclease" are used interchangeably to refer to an enzyme which possesses endonucleolytic catalytic activity for polynucleotide cleavage.

As used herein, "expression system" is a system comprising at least one nucleic acid encoding a guided nucleic acid binding agent and at least one guide.

As used herein, "exogenous mRNA" means any mRNA that is introduced into an organism or cell and that is not synthesized by the recipient organism or cell itself. An exogenous mRNA can be isolated or purified from an organism or cell, can be transcribed *in vitro*, or can be produced by synthetic means. An exogenous mRNA comprises a coding region (e.g., an open reading frame (ORF)) encoding a polypeptide sequence.

As used herein, "fragment" when used in relation to a polypeptide or nucleic acid, means a polypeptide or nucleic acid sequence that is at least one amino acid shorter than a reference sequence but otherwise identical to the reference sequence.

As used herein, "gene editing", "genome editing", or "genomic editing" means a process of changing the nucleobase sequence of a genome (e.g., insertions, deletions, mutations, or changes in a nucleobase, including epigenetic status (e.g., methylation states for A or G)), either directly or through innate cellular processes, e.g., following the introduction of a double-stranded break or a nick. In certain embodiments, gene editing is mediated by a complex comprising a guided nucleic acid binding agent and a guide. In certain

embodiments, gene editing is mediated by a complex comprising a Cas protein, a dead Cas protein, or Cas fusion protein and a guide.

As used herein, "guide" means an oligonucleotide (a "single guide" or a "guide oligonucleotide") or a complex consisting of two or more oligonucleotides that are partially hybridized to one another (e.g. a "dual guide"), in both cases comprising a "target-recognition region" or a "spacer" and a "protein-recognition region", or a "scaffold". In embodiments in which a guide is a single guide, the target-recognition region or spacer and the protein-recognition region or scaffold are regions of the one oligonucleotide that constitutes the single guide. In embodiments in which a guide is a dual guide, the target-recognition region or spacer is a region of a first oligonucleotide, and the protein-recognition region is a portion of the complex that includes regions of each of the two oligonucleotides. In certain embodiments, a guide directs a guided nucleic acid binding agent to a target sequence of a target DNA.

As used herein, "guided nucleic acid binding agent" means a polypeptide comprising (1) a region that interacts with the protein-recognition region of a guide (or "scaffold"); and (2) a region that interacts with the PAM of a target DNA. In certain embodiments, the target-recognition region or spacer of the guide causes the guided nucleic acid binding agent to specifically interact with a target DNA. A guided nucleic acid binding agent may comprise one or more active or inactive nuclease domains. A guided nucleic acid binding agent may comprise a heterologous domain. Guided nucleic acid binding agents include Cas enzymes, dead Cas proteins, and Cas fusion proteins.

As used herein, "hybridization", "hybridizing" or "hybridize" means the act or process of two complementary regions of strands of linked oligomeric subunits (*e.g.*, oligonucleotides, nucleic acids) annealing together to form a double-stranded region. While not limited to a particular mechanism, the most common mechanism of hybridization involves hydrogen bonding, which may be Watson-Crick, Hoogsteen, or reversed Hoogsteen hydrogen bonding, between complementary nucleobases.

As used herein, "identity," or "percent identity", with regard to an amino acid sequence, means the percentage of amino acids that are identical between two amino acid sequences when the amino acid sequences are aligned for maximal similarity. Identity measures the percent of identical matches between the smaller of two or more sequences when the sequences are aligned for maximal similarity. Identity of related polypeptides or nucleic acids can be calculated by known methods. "Percent (%) identity" refers to the percentage of amino acids or nucleobases in a candidate amino acid or nucleic acid sequence that are identical with the amino acids or nucleobases in the second sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent identity. Identity depends on a calculation of percent identity but may differ in value due to gaps and penalties introduced in the calculation. Methods and computer programs for sequence alignment are well known in the art, and include the BLAST suite (Stephen F. Altschul, et al (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search

programs", Nucleic Acids Res. 25:3389-3402), the Smith-Waterman algorithm (Smith, T.F. & Waterman, M.S. (1981) "Identification of common molecular subsequences." J. Mol. Biol. 147:195-197), the Needleman–Wunsch algorithm (Needleman, S.B. & Wunsch, C.D. (1970) "A general method applicable to the search for similarities in the amino acid sequences of two proteins." J. Mol. Biol. 48:443-453) and the Fast Optimal Global Sequence Alignment Algorithm (FOGSAA) for global alignment of amino acid or nucleic acid sequences.

As used herein, the term "interact with", with regards to two macromolecules (e.g. a polypeptide and an oligonucleotide; two polypeptides with each other; or a polypeptide and a nucleic acid), means that multiple non-covalent stabilizing interactions form between the two macromolecules, including but not limited to dipolar interactions (hydrogen bonds), cation-π or π-π stacking interactions, electrostatic interactions (salt bridges), and hydrophobic interactions.

As used herein, the term "internucleoside linkage" is the covalent linkage between adjacent nucleosides in an oligonucleotide. As used herein "modified internucleoside linkage" means any internucleoside linkage other than a phosphodiester internucleoside linkage.

As used herein, "lipid" refers to a group of organic compounds that include, but are not limited to, esters of fatty acids and are generally characterized by being poorly soluble in water, but soluble in many organic solvents.

As used herein, "lipid nanoparticle" is used interchangeably with "LNP" and in some embodiments refers to a composition comprising an ionizable and/or a cationic lipid, a non-cationic lipid, a sterol and a polymer lipid. In some embodiments, the non-cationic lipid is a neutral or zwitterionic lipid. In some embodiments, the sterol is cholesterol. In some embodiments, the polymer lipid is a pegylated ("PEG") lipid.

As used herein, "mismatch" or "non-complementary" means a nucleobase of a first oligonucleotide that is not complementary with the corresponding nucleobase of a second oligonucleotide or target nucleic acid when the first and second oligonucleotide are aligned.

A "modified guide" means a guide comprising a modified internucleoside linkage, a modified sugar moiety, and/or a modified nucleobase.

As used herein, "modulating" refers to changing or adjusting a feature in a cell, tissue, organ or organism.

As used herein, "motif" means the pattern of unmodified and/or modified sugar moieties, nucleobases, and/or internucleoside linkages, in an oligonucleotide.

As used herein, "messenger RNA" ("mRNA") is any ribonucleic acid (RNA) or modified ribonucleic acid that encodes at least one protein, including naturally-occurring, non-naturally-occurring, or modified

polymers of amino acids, and can be translated to produce the encoded protein *in vitro*, *in vivo*, *in situ*, or *ex vivo*. An mRNA may contain one or more modified nucleotides.

As used herein, "natural amino acid" means Gly or the $_{L}$-isomer of each of the following: Ala, Arg, Asn, Asp, Cys, Gln, Glu, His, Ile, Lys, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val.

As used herein, "nick" refers to a break in one strand of a double-stranded DNA.

As used herein, "non-cationic lipid" is a lipid molecule that carries a nonexchangeable net charge that is not positive. A "non-cationic lipid" can be a neutral, zwitterionic or negatively charged lipid. A "non-cationic lipid" may also be referred to as a helper lipid or a structural lipid.

As used herein, "non-natural amino acid" means any amino acid other than the standard twenty amino acids encoded by the human genetic code, including $_{D}$-isomers of each of the following: Ala, Arg, Asn, Asp, Cys, Gln, Glu, His, Ile, Lys, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val. A non-natural amino acid may have a modified or functionalized side chain (e.g., through the attachment of a linker) Examples of non-natural amino acids include, but are not limited to, allo-isoleucine, 2-amino-3-ethyl-pentanoic acid, aminoisobutyric acid, aminobutyric acid, azetidine, 7-azatryptophan, 6-azidolysine, β-cyclobutylalanine, β-methyl isoleucine, 4, 4-biphenylalanine, cis-hydroxyproline, cyclobutyl glycine, cyclohexyl glycine, cyclopentyl alanine, cyclopentyl glycine, 2,6-dimethyl tyrosine, 3,3-diphenyl alanine, 4-trans-hydroxy-L-proline, 1-napthaylalanine, 2-napthylalanine, N-methyl alanine, 1-methyl histidine, 3-methyl histidine, N-methyl-tryptophan, pipecolic acid,4-pyridylalanine, sarcosine, t-butyl alanine, or 3-t-butyl tyrosine.

As used herein, "nucleobase" means an unmodified nucleobase or a modified nucleobase. A nucleobase is a heterocyclic moiety. As used herein an "unmodified nucleobase" is adenine (A), thymine (T), cytosine (C), uracil (U), or guanine (G). As used herein, a "modified nucleobase" is a group of atoms other than unmodified A, T, C, U, or G capable of pairing with at least one other nucleobase. A "5-methyl cytosine" is a modified nucleobase. A universal base is a modified nucleobase that can pair with any one of the five unmodified nucleobases.

As used herein, "nucleobase sequence" means the order of contiguous nucleobases in a nucleic acid or oligonucleotide independent of any sugar or internucleoside linkage modification.

As used herein, "the nucleobase sequence of" or "the sequence of" a reference nucleobase SEQ ID NO, refers only to the nucleobase sequence provided in such SEQ ID NO and therefore, unless otherwise indicated, includes compounds wherein each sugar moiety and each internucleoside linkage, independently, may be modified or unmodified, irrespective of the presence or absence of modifications, indicated in the referenced SEQ ID NO.

As used herein, "nucleoside" means a compound or fragment of a compound comprising a nucleobase and a sugar moiety. The nucleobase and sugar moiety are each, independently, unmodified or modified.

As used herein, "oligonucleotide" means a strand of linked nucleosides connected via internucleoside linkages, wherein each nucleoside and internucleoside linkage independently may be modified or unmodified. Unless otherwise indicated, oligonucleotides consist of 8-150 linked nucleosides. As used herein, "modified oligonucleotide" means an oligonucleotide, wherein at least one nucleoside or internucleoside linkage is modified. As used herein, "unmodified oligonucleotide" means an oligonucleotide that does not comprise any nucleoside modifications or internucleoside modifications.

As used herein, "PAM interacting domain" or "PI domain" means a domain of a guided nucleic acid binding agent that can bind to or associate with one or more PAM sequences in the non-complementary strand of a target DNA. The PAM-interacting domain may confer specificity to one or more PAM sequences.

As used herein, "PAM sequence", "PAM", or "protospacer adjacent motif" refers to 2-9 linked nucleosides within the non-complementary strand of a target DNA that interacts with a guided nucleic acid binding agent. The PAM sequence may be adjacent to the 5'-end of a protospacer sequence. The PAM sequence may be adjacent to the 3'-end of a protospacer sequence.

As used herein, "the peptide sequence of" or "the polypeptide sequence of" or "the sequence of" a reference peptide/polypeptide SEQ ID NO. refers to the linear, amide-bond-linked amino acid sequence provided in such SEQ ID NO., even in cases where the given peptide or polypeptide contains one or more modified side chains that that link to another moiety.

As used herein, "pharmaceutically acceptable carrier or diluent" means any substance suitable for use in administering to a subject. Certain such carriers enable pharmaceutical compositions to be formulated as, for example, tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspension, and lozenges for the oral ingestion by a subject. In certain embodiments, a pharmaceutically acceptable carrier or diluent is sterile water, distilled water for injection, sterile saline, sterile buffer solution or sterile artificial cerebrospinal fluid.

As used herein "pharmaceutically acceptable salt(s)" means physiologically and pharmaceutically acceptable salt(s) of oligomeric compounds. Pharmaceutically acceptable salts retain the desired biological activity of the parent compound. The term "pharmaceutically acceptable salts" includes both acid and base addition salts. Pharmaceutically acceptable salts include those obtained by reacting the active compound functioning as a base. Those skilled in the art will further recognize that acid addition salts may be prepared by reaction of the compounds with the appropriate inorganic or organic acid via any of a number of known methods.

As used herein "pharmaceutical composition" means a mixture of substances suitable for administering to a subject. For example, in one embodiment a pharmaceutical composition may comprise an oligomeric agent and a sterile aqueous solution. In certain embodiments, a pharmaceutical composition shows activity in free uptake assay in certain cell lines. In some embodiments, pharmaceutical compositions include tablets, capsules, gel capsules, syrups, liquids, gels, powders, suspensions, solid dispersions, or combinations thereof. In some embodiments, a pharmaceutical composition may comprise an LNP.

As used herein, "protein-recognition region" means a portion of a guide that interacts with a guided nucleic acid binding agent. A "protein-recognition region" may be referred to as a "scaffold". A "scaffold" may adopt a tertiary structure that interacts with a Cas protein. In certain embodiments, a "protein-recognition region" comprises or consists of a "direct repeat region". In certain embodiments, a "protein-recognition region" consists of a "direct repeat region" and a "5'-stablizing region".

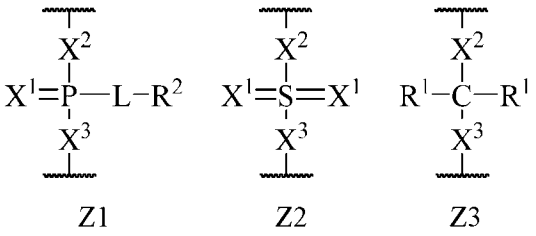
As used herein, a "protonatable lipid" is used interchangeably with an "ionizable lipid", being a lipid compound substantially protonated (e.g. it becomes 'cationic') at or below physiological pH (e.g., pH 7-7.5, or pH 7.4). A protonatable lipid containing an amine (for example, a tertiary amine), is a "protonatable amino lipid".

As used herein, the "protospacer sequence" of the target DNA refers to the reverse complement of the "target sequence" or "spacer" and is found in the non-complementary strand of the target DNA. In certain embodiments, the sequence of the target-recognition region (or "spacer") of a guide has at least 90%, 95%, 98%, or 99% identity to the protospacer sequence of the target DNA.

As used herein, "ribonucleoprotein" ("RNP") refers to a complex comprising a guide and a guided nucleic acid binding protein.

As used herein, "RuvC domain" or "RuvC" means a cation-dependent endonuclease domain of a protein with an active site comprising an aspartic acid (D) catalytic residue. In certain embodiments, the active site of a RuvC domain comprises aspartic acid (D) and glutamic acid (E) catalytic residues. In certain embodiments, a RuvC domain cleaves the non-complementary strand of a target DNA. In certain embodiments, a RuvC domain cleaves both strands of a target DNA. A RuvC domain may be formed from non-contiguous amino acids; for example, the RuvC domain of SpCas9 is split into RuvCI, RuvCII, and RuvCIII portions. A RuvC domain may have one or more substitutions that render it catalytically inactive. As used herein, a "stabilizing region" means a region of an oligonucleotide comprising 8-80 nucleotides that form a secondary structure. A "5'-stabilizing region" is appended at the 5'-end of an oligonucleotide, and a "3'-stabilizing region is appended at the 3'-end of an oligonucleotide.

As used herein, a "standard length internucleoside linkage" refers to an internucleoside linkage that has a structure represented by Formula Z1, Z2, or Z3:



Z1 Z2 Z3

wherein independently for each internucleoside linking group of Formula Z1, Z2, or Z3:

each X^1 is independently selected from O and S;

X^2 is selected from O, NR^1, CH_2, and S;

X^3 is selected from O, NR^1, CH_2, and S;

L is absent, NR^1, $N(R^1)SO_2$, –N=, O, C_1-C_6 alkylene, or C_1-C_6 heteroalkylene;

each R^1 is independently selected from H, C_1-C_6 alkyl, and substituted C_1-C_6 alkyl, or two R^1 on the same atom together form =O; and

R^2 is selected from -OH, -SH, C_1-C_{22} alkyl, substituted C_1-C_{22} alkyl, C_2-C_{22} alkenyl, substituted C_2-C_{22} alkenyl, cycloalkyl, substituted cycloalkyl, heterocyclyl, substituted heterocyclyl, heteroaryl, substituted heteroaryl, aryl, and substituted aryl;

wherein when a group is substituted, it comprises one or more substituent groups selected from halo, -OH, $-N(R^1)_2$, -O-C_1-C_6 alkyl, C_1-C_{22} alkyl, C_2-C_{22} alkenyl, cycloalkyl, heterocyclyl, heteroaryl, and aryl.

As used herein, “subject” refers to a human or non-human animal, including, but not limited to, mice, rats, rabbits, dogs, cats, pigs, and non-human primates, including, but not limited to, monkeys and chimpanzees.

As used herein, “sugar moiety” means an unmodified sugar moiety or a modified sugar moiety. As used herein, “unmodified sugar moiety” means a 2’-OH(H) ribosyl moiety, as found in RNA (an “unmodified RNA sugar moiety”), or a 2’-H(H) deoxyribosyl sugar moiety, as found in DNA (an “unmodified DNA sugar moiety”). Unmodified sugar moieties have one hydrogen at each of the 1’, 3’, and 4’ positions, an oxygen at the 3’ position, and two hydrogens at the 5’ position. As used herein, “modified sugar moiety” or “modified sugar” means a modified furanosyl sugar moiety or a sugar surrogate.

As used herein, "sugar surrogate" means a modified sugar moiety having other than a furanosyl moiety that can link a nucleobase to an internucleoside linkage. Modified nucleosides comprising sugar

surrogates can be incorporated into one or more positions within an oligonucleotide and such oligonucleotides are capable of hybridizing to complementary oligomeric compounds or target nucleic acids.

As used herein, "target-recognition region" refers to a 12-30 linked nucleoside portion of a guide that is complementary to the "target sequence" within the complementary strand of a target DNA. A target-recognition region may be fully complementary to the target sequence within the complementary strand of the target DNA. A "target-recognition region" may be referred to as a "spacer".

As used herein, "target sequence" refers to the 12-30 linked nucleoside portion of a nucleic target that is complementary to the "target-recognition region" of a guide. The target sequence is within the complementary strand of a target DNA.

As used herein, "treating" means improving a subject's disease or condition by administering a composition, an LNP, a guided nucleic acid binding agent, a guide, an editing system, a nucleic acid, an expression system or a viral vector as herein disclosed. In certain embodiments, treating a subject improves a symptom relative to the same symptom in the absence of the treatment. In certain embodiments, treatment reduces in the severity or frequency of a symptom, or delays the onset of a symptom, slows the progression of a symptom, or slows the severity or frequency of a symptom.

As used herein, "therapeutically effective amount" means an amount of an oligomeric agent or pharmaceutical composition that provides a therapeutic benefit to a subject. For example, a therapeutically effective amount improves a symptom of a disease.

As used herein, "alkyl" refers to a saturated straight or branched hydrocarbon substituent group containing up to twenty four carbon atoms. Examples of alkyl groups include without limitation, methyl, ethyl, propyl, butyl, isopropyl, n-hexyl, octyl, decyl, dodecyl and the like. Alkyl groups typically include from 1 to 20 carbon atoms ("C_1-C_{20} alkyl"), more typically from 1 to 12 carbon atoms ("C_1-C_{12} alkyl") with from 1 to 6 carbon atoms ("C_1-C_6 alkyl") being more preferred. Alkyl groups as used herein may optionally include one or more further substituent groups.

As used herein, "alkylene" refers to a fully saturated, straight or branched divalent hydrocarbon chain radical, and typically having from one to twelve or more carbon atoms. Non-limiting examples of C_1-C_{12} alkylene include methylene, ethylene, propylene, n-butylene, ethenylene, propenylene, n-butenylene, propynylene, n-butynylene, and the like. The alkylene chain is attached to the rest of the molecule through a single bond and to the radical group through a single bond. The points of attachment of the alkylene chain to the rest of the molecule and to the radical group can be through one carbon or any two carbons within the chain. Unless stated otherwise specifically, an alkylene chain can be optionally substituted.

As used herein, "alkenyl," refers to a straight or branched hydrocarbon chain substituent group containing up to twenty four carbon atoms and having at least one carbon-carbon double bond. Examples of

alkenyl groups include without limitation, ethenyl, propenyl, butenyl, 1-methyl-2-buten-1-yl, dienes such as 1,3-butadiene and the like. Alkenyl groups typically include from 2 to 20 carbon atoms, more typically from 2 to 12 carbon atoms with from 2 to 6 carbon atoms being more preferred. Alkenyl groups as used herein may optionally include one or more further substituent groups.

As used herein, "alkynyl", refers to a straight or branched hydrocarbon substituent group containing up to twenty four carbon atoms and having at least one carbon-carbon triple bond. Examples of alkynyl groups include, without limitation, ethynyl, 1-propynyl, 1-butynyl, and the like. Alkynyl groups typically include from 2 to 20 carbon atoms, more typically from 2 to 12 carbon atoms with from 2 to 6 carbon atoms being more preferred. Alkynyl groups as used herein may optionally include one or more further substituent groups.

As used herein, "alkoxy" refers to an alkyl-O- substituent group, where alkyl is as defined herein. Examples of alkoxy groups include without limitation, methoxy, ethoxy, propoxy, isopropoxy, n-butoxy, sec-butoxy, tert-butoxy, n-pentoxy, neopentoxy, n-hexoxy and the like. Alkoxy groups as used herein may optionally include further substituent groups.

As used herein, “amino” refers to the $-NH_2$ radical.

As used herein, "aryl" refers to a carbocyclic ring system substituent group having one or more aromatic rings. The aryl may be monocyclic or may include two or more fused rings. Examples of aryl groups include without limitation, phenyl, naphthyl, tetrahydronaphthyl, indanyl, indenyl and the like. Preferred aryl ring systems have from 6 to 10 ring atoms. Aryl groups as used herein may optionally include further substituent groups.

As used herein, “aralkyl” or “arylalkyl” refers to a radical of the formula $-R_b-R_c$ where R_b is an alkylene group as defined herein and R_c is one or more aryl radicals as defined herein, for example, benzyl, diphenylmethyl and the like. Unless stated otherwise specifically, an aralkyl group can be optionally substituted.

As used herein, “azido” refers to the $-N_3$ group.

As used herein, “cyano” refers to the –CN radical.

As used herein, "cycloalkyl" refers to a saturated or unsaturated carbocyclic ring system substituent group that does not include an aromatic ring. The cycloalkyl may be monocyclic or may include two or more fused rings. Examples of cycloalkyl groups include without limitation, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cyclohexenyl, and the like. Preferred cycloalkyl ring systems have from 3 to 10 ring atoms (“C_3-C_{10} cycloalkyl”). Cycloalkyl groups as used herein may optionally include further substituent groups.

As used herein, "haloalkyl" refers to an alkyls, as defined herein, that is substituted by one or more halos, as defined herein, e.g., trifluoromethyl, difluoromethyl, trichloromethyl, 2,2,2-trifluoroethyl, 1,2-difluoroethyl, 3-bromo-2-fluoropropyl, 1,2-dibromoethyl, and the like. Unless stated otherwise specifically, a haloalkyl group can be optionally substituted.

As used herein, "halo" or "halogen" refers to a substituent group selected from fluoride, chloride, bromide and iodide.

As used herein, "heteroaryl" refers to a substituent group comprising a ring system in which at least one of the rings is aromatic, and at least one ring includes one or more ring heteroatoms. The heteroaryl may be monocyclic or may include two or more fused rings. Heteroaryl groups include at least one ring atom selected from sulfur, nitrogen or oxygen, wherein the sulfur is optionally present as a sulfoxide or sulfone, and wherein the nitrogen is optionally present as an N-oxide. Examples of heteroaryl groups include without limitation, pyridinyl, pyrazinyl, pyrimidinyl, pyrrolyl, pyrazolyl, imidazolyl, thiazolyl, oxazolyl, isooxazolyl, thiadiazolyl, thiophenyl, furanyl, quinolinyl, and the like. Heteroaryl groups as used herein may optionally include further substituent groups.

As used herein, "heteroalkylene" refers to an alkylene radical where one, two or three carbons in the alkylene chain is replaced by -O-, N(H, alkyl, or substituted alkyl), S, SO, SO_2, or CO.

As used herein, "heterocyclyl" refers to a substituent group comprising a ring system in which none of the rings are aromatic, and at least one ring includes one or more ring heteroatoms. Heterocyclyl is also meant to include fused ring systems including systems where one or more of the fused rings contain no heteroatoms. Heterocyclyl groups include at least one ring atom selected from sulfur, nitrogen or oxygen, wherein the sulfur is optionally present as a sulfoxide or sulfone. Examples of heterocyclyl groups include without limitation, morpholino, oxirane, tetrahydropyranyl, tetrahydrothienyl, sulfolanyl, and the like. Heterocyclyl groups as used herein may optionally include further substituent groups.

As used herein, "hydroxy" or "hydroxyl" refers to the –OH radical.

As used herein, "imino" refers to the =NH substituent.

As used herein, "ring" refers to a cyclic group which can be fully saturated, partially saturated, or fully unsaturated. A ring can be monocyclic, bicyclic, tricyclic, or tetracyclic. Unless stated otherwise specifically, a ring can be optionally substituted.

The term "substituted", with respect to chemical groups means, unless otherwise indicated, a group is substituted with 1, 2, 3, 4, or 5 or more substituent groups selected from halo (e.g., perhalo), hydroxy, azido, SH, CN, OCN, nitro, C_1-C_{20} alkyl (e.g., C_1-C_2 alkyl), C_1-C_{10} substituted alkyl (e.g., CF_3), C_2-C_{10} alkenyl, C_2-C_{10} alkynyl, C_1-C_{10} heteroalkyl, C_1-C_{10} alkoxy, C_1-C_{10} substituted alkoxy (e.g., OCF_3), S-alkyl, N(R_m)-alkyl,

O-alkenyl, S-alkenyl, N(R_m)-alkenyl, O-alkynyl, S-alkynyl, N(R_m)-alkynyl, O-alkylenyl-O-alkyl, aralkyl, O-aralkyl, C_{3-10}cycloalkyl, C_{6-10}aryl, heterocyclyl, heteroaryl, N(R_m)(R_n), C(O)N(R_m)(R_n), N(R_m)(R_n)C(O)R_m, $S(O)_2N(R_m)(R_n)$, N(R_m)(R_n)$S(O)_2R_m$, OC(O)N(R_m)(R_n), N(R_m)C(O)N(R_m)(R_n), N(R_m)C(O)OR_n, C(O)OR_m, and OC(O)R_m, where each R_m and R_n is, independently, H, OH, an amino protecting group, or substituted or unsubstituted C_1-C_{10} alkyl. Substituent groups of this paragraph can be unsubstituted or further substituted at a carbon atom with one or more groups independently selected from: hydroxyl, amino, alkoxy, carboxy, benzyl, phenyl, cyano, nitro (NO_2), thiol, thioalkoxy, thioalkyl, halogen, alkyl, aryl, alkenyl and alkynyl.

As used herein, "thioalkyl" refers to a radical of the formula -SR_a where R_a is an alkyl, alkenyl, or alkynyl as defined herein typically containing one to twelve or more carbon atoms. Unless stated otherwise specifically, a thioalkyl group can be optionally substituted.

CERTAIN EMBODIMENTS

The present disclosure provides the following non-limiting numbered embodiments:

Embodiment 1. A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids, wherein the amino sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length portion of any of SEQ ID NOs: 4-95.

Embodiment 2. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 4-95.

Embodiment 3. A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids, wherein the amino sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length portion of any of SEQ ID NOs: 18, 20-24, 26-32, 34, 36-39, 41, 42, 44, 45, 50, 53-58, 75-76, 84, 85, 87-91, or 95.

Embodiment 4. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 18, 20-24, 26-32, 34, 36-39, 41, 42, 44, 45, 50, 53-58, 75-76, 84, 85, 87-91, or 95.

Embodiment 5. The guided nucleic acid binding agent of any of embodiments 1-4, consisting of the first polypeptide.

Embodiment 6. The guided nucleic acid binding agent of any of embodiments 1-5, wherein the first polypeptide is a Cas polypeptide.

Embodiment 7. The guided nucleic acid binding agent of any of embodiments 1-6, wherein the first polypeptide comprises a RuvC domain.

Embodiment 8. The guided nucleic acid binding agent of any of embodiments 1-7, wherein the first polypeptide has exactly one catalytically active nuclease site.

Embodiment 9. The guided nucleic acid binding agent of embodiment 8, wherein the first polypeptide can introduce a double-stranded break in DNA.

Embodiment 10. The guided nucleic acid binding agent of any of embodiments 1-7, wherein the first polypeptide has zero catalytically active nuclease sites.

Embodiment 11. The guided nucleic acid binding polypeptide of embodiment 10, wherein the first polypeptide comprises a RuvC domain and the RuvC domain contains an inactivating substitution of a conserved Asp.

Embodiment 12. The guided nucleic acid binding agent of any preceding embodiment, comprising a second polypeptide.

Embodiment 13. The guided nucleic acid binding agent of embodiment 12, wherein the second polypeptide comprises a heterologous domain.

Embodiment 14. The guided nucleic acid binding agent of embodiment 13, wherein the heterologous domain is selected from a transcriptional activator, a transcriptional repressor, a methyltransferase, a demethylase, a deaminase, an acetyltransferase, or a deacetylase.

Embodiment 15. The guided nucleic acid binding agent of any of embodiments 12-14, wherein the first polypeptide and the second polypeptide are fused to form a single protein.

Embodiment 16. The guided nucleic acid binding agent of embodiment 15, wherein the second polypeptide is fused to the N-terminus of the first polypeptide, forming a fusion protein.

Embodiment 17. The guided nucleic acid binding agent of embodiment 16, wherein the second polypeptide is fused to the C-terminus of the first polypeptide, forming a fusion protein.

Embodiment 18. The guided nucleic acid binding agent of any of any of embodiments 12-14, further comprising one or more additional fused heterologous domains.

Embodiment 19. The guided nucleic acid binding agent of embodiments 15-18, wherein the fusion protein is an epigenetic editing protein.

Embodiment 20. A guide, wherein the guide comprises a protein-recognition element, wherein the protein-recognition element binds to the first polypeptide of any of embodiments 1-19.

Embodiment 21. The guide of embodiment 20, wherein the guide is a single guide comprising an oligonucleotide, and the protein-recognition element is a protein-recognition region of the oligonucleotide.

Embodiment 22. The guide of embodiment 21, wherein the nucleobase sequence of the protein-recognition region of the guide has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 280-372.

Embodiment 23. The guide any of embodiments 20-22, wherein the guide comprises a modified oligonucleotide.

Embodiment 24. The guide of embodiment 23, wherein the guide comprises at least one modified sugar moiety.

Embodiment 25. The guide of embodiment 24, wherein the modified sugar moiety is selected from a 2'-OMe and a 2'-F.

Embodiment 26. The guide of embodiment 25, wherein the guide comprises a 2'-OMe or 2'-F sugar moiety within the first five nucleosides of the 5' end or the last five nucleosides 3' end of the guide.

Embodiment 27. The guide of any of embodiments 23-26, wherein the guide comprises a modified internucleoside linkage.

Embodiment 28. The guide of embodiment 27, wherein the modified internucleoside linkage is a phosphorothioate internucleoside linkage.

Embodiment 29. The guide of embodiment 27 or 28, wherein the guide comprises at least one modified internucleoside linkage within the first five nucleosides of the 5' end or the last five nucleosides at the 3' end of the guide.

Embodiment 30. The guide of any of embodiments 20-29, wherein the guide consists of an oligonucleotide.

Embodiment 31. The guide of any of embodiments 20-30, wherein the guide comprises a target-recognition element that is at least 90%, at least 95%, or 100% complementary to a target sequence.

Embodiment 32. An editing system comprising the nucleic acid binding agent of any of embodiments 1-19 and a guide of any of embodiments 20-31.

Embodiment 33. An editing system comprising:

a. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 4-95; and

b. A guide, wherein the nucleobase sequence of the protein-recognition region of the guide has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 280-372;

wherein the SEQ ID NO: of the first polypeptide and the SEQ ID NO: of the guide are selected from the same row of Table 5.

Embodiment 34. The editing system of embodiment 33, wherein the guided nucleic acid binding agent further comprises a second polypeptide that comprises at least one heterologous domain.

Embodiment 35. A nucleic acid encoding the first polypeptide of any of embodiments 1-19.

Embodiment 36. An expression system comprising:

a. a nucleic acid of embodiment 35;

b. a guide of any of embodiments 20-31.

Embodiment 37. An expression system comprising:

a. a nucleic acid of embodiment 35; and

b. a nucleic acid encoding the guide of any of embodiments 20-22.

Embodiment 38. The nucleic acid of embodiment 35 or expression system of embodiment 36 or 37, wherein the nucleic acid encoding the first polypeptide is an exogenous mRNA.

Embodiment 39. The nucleic acid of embodiment 35 or expression system of embodiment 36 or 37, wherein the nucleic acid encoding the first polypeptide is a DNA.

Embodiment 40. The expression system of any of embodiments 36-38, wherein the nucleic acid encoding the guide is an unmodified RNA.

Embodiment 41. The expression system of any of embodiments 36-38, wherein the nucleic acid encoding the guide is a DNA.

Embodiment 42. The expression system of embodiment 37, wherein the polypeptide and the guide are encoded by the same nucleic acid.

Embodiment 43. An LNP at least partially encapsulating the editing system of any of embodiments 32-34, the nucleic acid of embodiment 35, or the expression system of any of embodiments 36-42.

Embodiment 44. A viral vector comprising the nucleic acid of embodiment 35 or the expression system of any of embodiments 37-42.

Embodiment 45. A composition comprising:

a. the viral vector of embodiment 44;

b. a guide;

c. an LNP

Embodiment 46. A method of editing a target nucleic acid, comprising administering an LNP of embodiment 43, a viral vector of embodiment 44, or a composition of embodiment 45 to a subject.

Embodiment 47. A method of creating a discontinuity in a target nucleic acid, comprising administering an LNP of embodiment 43, a viral vector of embodiment 44, or a composition of embodiment 45 to a subject.

Embodiment 48. A method of gene silencing, comprising administering an LNP of embodiment 43, a viral vector of embodiment 44, or a composition of embodiment 45 to a subject.

Embodiment 49. A method of editing a target nucleic acid, comprising contacting a cell with an LNP of embodiment 43, a viral vector of embodiment 44, or a composition of embodiment 45.

Embodiment 50. A method of creating a double stranded break or a nick in a target DNA, comprising contacting a cell with an LNP of embodiment 43, a viral vector of embodiment 44, or a composition of embodiment 45.

Embodiment 51. The method of any of embodiments 46-49, wherein the target nucleic acid is a target DNA.

Embodiment 52. The method of any of embodiments 50-51, wherein the non-complementary strand of the target DNA comprises a sequence selected from Table 6 within 18-20 nucleobases from the cut site in the non-complementary strand.

Embodiment 53. A method of gene silencing, comprising contacting a cell with an LNP of embodiment 43, a viral vector of embodiment 44, or a composition of embodiment 45.

Embodiment 54. The method of any of embodiments 49-53, wherein the cell is in a subject.

Embodiment 55. The LNP of embodiment 43, viral vector of embodiment 44, or composition of embodiment 45 for use in therapy.

Embodiment 56. A complex comprising the editing system of embodiment 33 in contact with a target DNA, wherein the complementary strand of the DNA comprises a sequence complementary to the guide adjacent to a sequence complementary to a sequence selected from Table 6.

Embodiment 57. A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids, wherein the amino sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length portion of any of SEQ ID NOs: 4-95, 501-533, 600-601.

Embodiment 58. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 4-95, 501-533, 600-601.

Embodiment 59. A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids, wherein the amino sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length portion of any of SEQ ID NOs: 18, 20-24, 26-32, 34, 36-39, 41, 42, 44, 45, 50, 53-58, 75-76, 84, 85, 87-91, or 95.

Embodiment 60. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 18, 20-24, 26-32, 34, 36-39, 41, 42, 44, 45, 50, 53-58, 75-76, 84, 85, 87-91, or 95.

Embodiment 61. The guided nucleic acid binding agent of any of embodiments 57-60, consisting of the first polypeptide.

Embodiment 62. The guided nucleic acid binding agent of any of embodiments 57-61, wherein the first polypeptide is a Cas protein.

Embodiment 63. The guided nucleic acid binding agent of any of embodiments 57-62, wherein the first polypeptide comprises a RuvC domain.

Embodiment 64. The guided nucleic acid binding agent of any of embodiments 57-63, wherein the first polypeptide has exactly one catalytically active nuclease site.

Embodiment 65. The guided nucleic acid binding agent of embodiment 64, wherein the first polypeptide can introduce a double-stranded break in DNA.

Embodiment 66. The guided nucleic acid binding agent of any of embodiments 57-63, wherein the first polypeptide has zero catalytically active nuclease sites.

Embodiment 67. The guided nucleic acid binding polypeptide of embodiment 66, wherein first polypeptide comprises a RuvC domain and the RuvC domain contains an inactivating substitution of a conserved Asp.

Embodiment 68. The guided nucleic acid binding agent of any preceding embodiment, comprising a second polypeptide.

Embodiment 69. The guided nucleic acid binding agent of embodiment 68, wherein the second polypeptide comprises a heterologous domain.

Embodiment 70. The guided nucleic acid binding agent of embodiment 69, wherein the functional domain is selected from a transcriptional activator, a transcriptional repressor, a methyltransferase, a demethylase, a deaminase, an acetyltransferase, or a deacetylase.

Embodiment 71. The guided nucleic acid binding agent of any of embodiments 68-70, wherein the first polypeptide and the second polypeptide are fused to form a single protein.

Embodiment 72. The guided nucleic acid binding agent of embodiment 71, wherein the second polypeptide is fused to the N-terminus of the first polypeptide, forming a fusion protein.

Embodiment 73. The guided nucleic acid binding agent of embodiment 71, wherein the second polypeptide is fused to the C-terminus of the first polypeptide, forming a fusion protein.

Embodiment 74. The guided nucleic acid binding agent of any of any of embodiments 68-70, further comprising one or more additional fused heterologous domains.

Embodiment 75. The guided nucleic acid binding agent of embodiments 71-74, wherein the fusion protein is an epigenetic editing protein.

Embodiment 76. A guide, wherein the guide comprises a protein-recognition region, wherein the protein-recognition region binds to the first polypeptide of any of embodiments 57-75.

Embodiment 77. The guide of embodiment 76, wherein the guide is a single guide comprising an oligonucleotide, and the protein-recognition region comprises a direct repeat region.

Embodiment 78. The guide of embodiment 77, wherein the nucleobase sequence of the direct repeat region of the guide has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 1176-1228.

Embodiment 79. The guide any of embodiments 76-78, wherein the guide comprises a modified oligonucleotide.

Embodiment 80. The guide of embodiment 79, wherein the guide comprises at least one modified sugar moiety.

Embodiment 81. The guide of embodiment 80, wherein the modified sugar moiety is selected from a 2'-OMe and a 2'-F.

Embodiment 82. The guide of embodiment 81, wherein the guide comprises a 2'-OMe or 2'-F sugar moiety within the first five nucleosides of the 5' end or the last five nucleosides 3' end of the guide.

Embodiment 83. The guide of any of embodiments 79-82, wherein the guide comprises a modified internucleoside linkage.

Embodiment 84. The guide of embodiment 83, wherein the modified internucleoside linkage is a phosphorothioate internucleoside linkage.

Embodiment 85. The guide of embodiment 83 or 84, wherein the guide comprises at least one modified internucleoside linkage within the first five nucleosides of the 5' end or the last five nucleosides at the 3' end of the guide.

Embodiment 86. The guide of any of embodiments 76-85, wherein the guide consists of an oligonucleotide.

Embodiment 87. The guide of any of embodiments 76-86, wherein the guide comprises a target-recognition region that is at least 90%, at least 95%, or 100% complementary to a target sequence.

Embodiment 88. The guide of any of embodiments 76-87, wherein the guide comprises a stabilizing region.

Embodiment 89. The guide of embodiment 88, wherein the stabilizing region is at the 5'-end of the guide.

Embodiment 90. The guide of embodiment 88, wherein the stabilizing region is at the 3'-end of the guide.

Embodiment 91. The guide of any of embodiments 88-90, wherein the stabilizing region has a sequence that is 90%, 95%, or 100% identical to any of SEQ ID Nos: 488-494, 498, or 1172-1175.

Embodiment 92. An editing system comprising the nucleic acid binding agent of any of embodiments 57-75 and a guide of any of embodiments 76-91.

Embodiment 93. An editing system comprising:

a. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least

96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 4-95, 501-533, or 600-601; and

b. A guide, wherein the nucleobase sequence of the direct repeat region of the guide has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 1176-1228;

wherein the SEQ ID NO: of the first polypeptide and the SEQ ID NO: of the guide are selected from the same row of Table 52.

Embodiment 94. The editing system of embodiment 93, wherein the guided nucleic acid binding agent further comprises a second polypeptide that comprises at least one heterologous domain.

Embodiment 95. The editing system of embodiment 93 or 94, comprising a ribonucleoprotein (RNP), wherein the RNP comprises the guided nucleic acid binding agent and the guide.

Embodiment 96. A nucleic acid encoding the first polypeptide of any of embodiments 57-75.

Embodiment 97. An expression system comprising:

a. a nucleic acid of embodiment 96; and

b. a guide of any of embodiments 76-91.

Embodiment 98. An expression system comprising:

a. a nucleic acid of embodiment 96; and

b. a nucleic acid encoding the guide of any of embodiments 76-78.

Embodiment 99. The nucleic acid of embodiment 96 or expression system of embodiment 97 or 98, wherein the nucleic acid encoding the first polypeptide is an exogenous mRNA.

Embodiment 100. The nucleic acid of embodiment 96 or expression system of embodiment 97 or 98, wherein the nucleic acid encoding the first polypeptide is a DNA.

Embodiment 101. The expression system of any of embodiments 97-100, wherein the nucleic acid encoding the guide is an unmodified RNA.

Embodiment 102. The expression system of any of embodiments 97-100, wherein the nucleic acid encoding the guide is a DNA.

Embodiment 103. The expression system of embodiment 98, wherein the polypeptide and the guide are encoded by the same nucleic acid.

Embodiment 104. An LNP at least partially encapsulating the editing system of any of embodiments 93-95, the nucleic acid of any of embodiments 96, 99 or 100, or the expression system of any of embodiments 97, 98, or 101-103.

Embodiment 105. A viral vector comprising the nucleic acid of any of embodiments 96, 99 or 100, or the expression system of any of embodiments 97, 98, or 101-103.

Embodiment 106. A composition comprising:

the guided nucleic acid binding agent of any of embodiments 57-75 and the guide of any of embodiments 76-91;

the editing system of any of embodiments 92-95;

the expression system of any of embodiments 97, 98, or 101-103;

the LNP of embodiment 104; or

the viral vector of embodiment 105.

Embodiment 107. A method of editing a target nucleic acid, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 108. A method of creating a discontinuity in a target nucleic acid, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 109. A method of creating a double stranded break or a nick in a target DNA, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 110. A method of gene silencing, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 111. A method of gene activation, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 112. A method of homologous gene repair, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 113. A method of gene visualization, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 114. A method of epigenetic modulation of gene expression, comprising administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 115. A method of editing a target nucleic acid, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 116. A method of creating a discontinuity in a target nucleic acid, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 117. A method of creating a double stranded break or a nick in a target DNA, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 118. A method of gene silencing, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 119. A method of gene activation, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 120. A method of homologous gene repair, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 121. A method of gene visualization, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 122. A method of epigenetic modulation of gene expression, comprising contacting a cell with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 to a subject.

Embodiment 123. The method of any of embodiments 107, 108, 115 or 116, wherein the target nucleic acid is a target DNA.

Embodiment 124. The method of any of embodiments 109, 117 or 123, wherein the non-complementary strand of the target DNA comprises a sequence selected from Table 6 within 18-20 nucleobases from the cut site in the non-complementary strand.

Embodiment 125. The method of any of embodiments 115-122, wherein the cell is in a subject.

Embodiment 126. The LNP of embodiment 104, viral vector of embodiment 105, or composition of embodiment 106 for use in therapy.

Embodiment 127. A complex comprising the editing system of embodiment 93 in contact with a target DNA, wherein the complementary strand of the DNA comprises a sequence complementary to the guide adjacent to a sequence complementary to a sequence selected from Table 6.

Embodiment 128. A method of treating a disease or disorder in a subject, wherein the method comprises administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 129. A method of autologous cell therapy, wherein the method comprises providing an autologous cell, contacting an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 with the cell, and administering the cell to a subject.

Embodiment 130. A method of allogeneic cell therapy, wherein the method comprises providing an allogeneic cell, contacting an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 with the cell, and administering the cell to a subject.

Embodiment 131. A cell comprising an edited target nucleic acid, produced by the method of embodiment 115.

Embodiment 132. A cell comprising a discontinuity in a target nucleic acid, produced by the method of embodiment 116.

Embodiment 133. A cell comprising a double stranded break or a nick in a target DNA, produced by the method of embodiment 117.

Embodiment 134. A cell comprising a silenced gene, produced by the method of embodiment 118.

Embodiment 135. A cell comprising an activated gene, produced by the method of embodiment 119.

Embodiment 136. A cell comprising a homologously repaired gene, produced by the method of embodiment 120.

Embodiment 137. A cell comprising a gene that can be visualized, produced by the method of embodiment 121.

Embodiment 138. A gene, the expression of which has been epigenetically modified, produced by the method of embodiment 122.

Embodiment 139. Use of an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 in the manufacture of a medicament for treating a disease or disorder in a subject.

Embodiment 140. Use of a cell of any of embodiments 131-138 in the manufacture of a medicament for treating a disease or disorder in a subject.

Embodiment 141. Use of an autologous cell, wherein the cell has been contacted with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 in the manufacture of a medicament for autologous cell therapy.

Embodiment 142. Use of an allogeneic cell, wherein the cell has been contacted with an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106 in the manufacture of a medicament for allogeneic cell therapy.

Embodiment 143. The cell of any of embodiments 131-137 for use in therapy.

Embodiment 144. A method of diagnosing a disease or disorder in a subject, wherein the method comprises administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 145. A method of assessing responsiveness to a treatment of a disease or disorder in a subject, wherein the method comprises administering an LNP of embodiment 104, a viral vector of embodiment 105, or a composition of embodiment 106.

Embodiment 146. A kit comprising an LNP of embodiment 104, a viral vector of embodiment 105, a composition of embodiment 106, or a cell of any of embodiments 131-138 and optionally instructions for use, optionally wherein the kit is for, or when used for, treating a disease or disorder in a subject or a cell thereof.

Embodiment 147. A compound comprising an oligonucleotide consisting of 25 to 150 linked nucleosides, wherein the 5'-terminus of the oligonucleotide consists of 15 or 16 nucleotides having a

sequence with at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length region of SEQ ID NOs: 1172-1174.

Embodiment 148. The compound of embodiment 147, wherein the oligonucleotide comprises at least one modified nucleoside or at least one modified internucleoside linkage.

Embodiment 149. The compound of embodiment 147 or 148, wherein the compound comprises a direct repeat region.

Embodiment 150. The compound of any of embodiments 147-149, wherein the compound comprises a target-recognition region.

Embodiment 151. The compound of any of embodiments 147-150, wherein the target-recognition region is complementary to a mammalian target.

Embodiment 152. The compound of any of embodiments 147-151, wherein the compound is a guide, and wherein the gene editing activity of the guide is increased relative to the gene editing activity of an otherwise identical guide that is truncated by 15 or 16 nucleotides at the 5'-end compared to the compound of any of embodiments 147-151.

Embodiment 153. An editing system comprising the compound of any of embodiments 147-152.

Embodiments provided herein are directed to guides, guided nucleic acid binding agents, editing systems comprising guided nucleic acid binding agents and guides, expression systems comprising nucleic acids encoding guided nucleic acid binding agents and guides, compositions, methods of use thereof, cells produced by such methods, methods using such cells, and kits. In certain embodiments, the guided nucleic acid binding agent comprises or consists of a Cas protein. In certain embodiments, the guided nucleic acid binding agent comprises or consists of a Cas enzyme. In certain embodiments, the guided nucleic acid binding agent comprises a Cas protein comprising a RuvC domain. In certain embodiments, the guided nucleic acid binding agent comprises a Class 2, Type V Cas polypeptide. In certain embodiments, the guided nucleic acid binding agent comprises a Class 2, Type V Cas endonuclease. In certain embodiments, the guided nucleic acid binding agent comprises a Cas protein comprising a RuvC. In certain embodiments, the nucleic acid binding agent comprises a modified Cas protein with an inactivating mutation in the nuclease active site (a "dead Cas" or "dCas"). In certain embodiments, the protein is isolated. In certain embodiments, the protein is engineered.

In certain embodiments, an editing system comprises a guided nucleic acid binding agent and a guide. In certain embodiments, a guide consists of a single oligonucleotide and is a single guide. In certain embodiments, a guide comprises a target-recognition region and a protein-recognition region. In certain embodiments, the guide is a single guide consisting of an oligonucleotide that comprises a target-recognition region and a protein-recognition region. The target-recognition region is a region of an oligonucleotide that

has a sequence complementary to an equal length portion of a target sequence within the complementary strand of a target DNA.

I. Certain Nucleic Acid Binding Agents

A. Certain Polypeptides

In certain embodiments, a guided nucleic acid binding agent is an isolated or engineered polypeptide. In certain embodiments, the polypeptide consists of the Cas protein, or is a fusion protein comprising a Cas protein. In certain embodiments, the Cas protein is a Cas enzyme. In alternative embodiments, the Cas protein is a dead Cas protein. In certain embodiments, the polypeptide that constitutes a Cas protein has one or more domains. In certain embodiments, a domain of the polypeptide is a RuvC domain, which is comprised of one or more subdomains such as a RuvCI, RuvCII, and/or RuvCIII subdomains, e.g., as described herein. In certain embodiments, a domain of the polypeptide is a PAM-interacting (PI) domain, e.g., as described herein. In certain embodiments, the polypeptide comprises a RuvC domain. In certain embodiments, the polypeptide comprises a RuvC domain and a PI domain. In certain embodiments, the polypeptide comprises a RuvC domain, a PI domain, and a nuclear localization sequence (NLS). In certain embodiments, a RuvC domain is split into two or three noncontiguous subdomains, termed RuvCI, RuvCII, and RuvCIII. In certain embodiments, the RuvC domain is part of a NUC lobe.

In certain embodiments, the polypeptide further comprises a REC lobe. In certain embodiments, the REC lobe is made up of subdomains corresponding to two REC domains, RECI, and RECII. In certain embodiments, RECI is split into two noncontiguous subdomains RECIa and RECIb. In certain embodiments, the polypeptide further comprises a bridge helix domain.

In certain embodiments, the polypeptide consists of a Class 2, Type V Cas protein, or is a fusion protein comprising a Class 2, Type V Cas protein. In certain embodiments, the Class 2, Type V Cas protein is a Class 2, Type V Cas enzyme. In alternative embodiments, the Class 2, Type V Cas protein is a dead Class 2, Type V Cas protein. Although members of Class 2, Type V CRISPR-Cas systems have differences, there exist common characteristics that are distinguishable from Cas9 systems. For example, Class 2, Type V systems typically possess a single RNA-guided RuvC domain-containing effector but no HNH domain, and they recognize a T-rich motif PAM 5′ upstream to the target region on the non-targeted strand. This is different to Cas9 systems that rely on G-rich PAM at the 3′ end of target sequences. Type V systems also typically generate staggered double-stranded breaks distal to the PAM sequence, unlike Cas9, that generates a blunt end in the proximal site close to the PAM.

B. Protein Engineering

Methods of structure-guided protein engineering as well as protein engineering via directed evolution have been applied to a variety of Cas proteins (see Liu, et al., *Trends in Biotechnology,* 39(3):262-273, 2021).

1. Certain Heterologous Domains

In certain embodiments, a guided nucleic acid binding agent is a Cas fusion protein that includes a Cas protein and a heterologous domain. In certain embodiments, heterologous domain(s) from non-Cas proteins can be appended to a Cas protein, to add a function (see, e.g. WO2014/152432; WO2016/063264; WO2022/162247; WO2022/140577; US2019/0233805; WO2023/235818), resulting in a Cas fusion protein. In certain embodiments, the Cas protein portion of the fusion protein is used to target the heterologous domain(s) to a DNA target of interest. In certain embodiments, the Cas protein portion of the fusion protein is a dead Cas protein.

In certain such embodiments, the heterologous domains(s) are transcriptional activators or transcriptional repressors, or domains thereof. For example, in some cases, the heterologous domains(s) inhibit transcription (e.g., the heterologous domain is a transcriptional repressor, a polypeptide that functions via recruitment of transcription inhibitor proteins, a polypeptide that leads to modification of target DNA such as methylation, a polypeptide that leads to recruitment of a DNA modifier, a polypeptide that modulates histones associated with target DNA, or a polypeptide that recruits a histone modifier such as those that modify acetylation and/or methylation of histones). In some cases, the heterologous domain(s) increase transcription (e.g., the exogenous domain is a transcription activator, a polypeptide that acts via recruitment of transcription activator proteins, a polypeptide that leads to modification of target DNA such as demethylation, a polypeptide that leads to recruitment of a DNA modifier, a polypeptide that modulates histones associated with target DNA, or a polypeptide that recruits a histone modifier such as those that modify acetylation and/or methylation of histones).

In certain embodiments, a Cas fusion protein includes a heterologous domain that 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).

In certain embodiments, the heterologous domain(s) modify a second polypeptide (e.g. a histone) associated with a target nucleic acid (e.g., polypeptides having 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).

Examples of polypeptides that can be used to increase transcription include, but are not limited to, transcriptional activators such as VP16, VP64, VP48, VP160, MyoD1, HSF1, RTA, SET7/9, or domains thereof; p65 subdomain (e.g., from NFkB), and activation domains of EDLL and/or TAL (e.g., for activity in plants); histone lysine methyltransferases such as SET1A, SET1B, MLL1 to 5, ASH1, SYMD2, NSD1 or domains thereof; histone lysine demethylases such as JHDM2a/b, UTX, JMJD3 or domains thereof; histone acetyltransferases such as GCN5, PCAF, CBP, p300, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, SRC1, ACTR, P160, CLOCK or domains thereof; and DNA demethylases such as TenElevenTranslocation (TET) dioxygenase 1 (TET1CD), TET1, DME, DML1, DML2, ROS1, or domains thereof.

Examples of polypeptides that can be used to decrease transcription include, but are not limited to, transcriptional repressor polypeptides such as the Krüppel associated box (KRAB or SKD) domains; KOX1 repression domains; the Mad mSIN3 interaction domain (SID); the ERF repressor domain (ERD), the SRDX repression domain (e.g., for repression in plants); histone lysine methyltransferases such as Pr-SET7/8, SUV4-20H1, RIZ1, or domains thereof; histone lysine demethylases such as JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASCI, JMJD2D, JARID1A/RBP2, JARID1B/PLU-I, JARID1C/SMCX, JARID1D/SMCY; histone lysine deacetylases such as HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRTI, SIRT2, HDAC11, or domains thereof; 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, CMTI, CMT2 (plants), or domains thereof; and periphery recruitment elements such as Lamin A and Lamin B, or domains thereof.

In certain embodiments, the heterologous domain has enzymatic activity that modifies the target nucleic acid. Examples of enzymatic activity that can be provided by the heterologous domain include but are not limited to: nuclease activity such as that provided by a restriction enzyme (e.g., Fokl nuclease), methyltransferase activity such as that provided by a methyltransferase (e.g., Hhal DNA m5c-methyltransferase (M.Hhal), DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3a (DNMT3a), DNA methyltransferase 3b (DNMT3b), METI, DRM3 (plants), ZMET2, CMT1, CMT2 (plants)); demethylase activity such as that provided by a demethylase (e.g., Ten-Eleven Translocation (TET) dioxygenase 1 (TET1CD), TET1, DME, DML1, DML2, ROSI, 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 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), 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).

In certain embodiments, the heterologous domain 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). Examples of enzymatic activity (that modifies a protein associated

with a target nucleic acid) that can be provided by the heterologous domain 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 (SUV39Hl, also known as KMT1A), euchromatic histone lysine methyltransferase 2 (G9A, also known as KMT1C and EHMT2), SUV39H2, ESET/SETDBI, 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), 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), deacetylase activity such as that provided by a histone deacetylase (e.g., HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRTI, SIRT2, HDACll), 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.

Domains that have been appended to Cas proteins include FokI for highly specific genome editing (see, e.g. WO2014/152432), transcriptional repressor domains (e.g. KRAB; see, e.g., Alerasool, et al., *Nature Methods*, 1093-1096, 2020) and transcriptional activator domains (e.g. VP64 – see, e.g., WO2014/197748; p65; and RTA) for the regulation of gene expression, histone modification enzymes such as methyltransferases for epigenetic editing (see, e.g., WO2016/103233), and fluorescent proteins for genome imaging (see, e.g. *Trends in Biotechnology,* 39(3):262-273, 2021). One or more of these polypeptides can be appended to a Cas protein at either the N-terminus or the C-terminus of the Cas protein. In certain embodiments, multiple domains are appended to a Cas protein at the C-terminal, the N-terminal, or both (see, e.g. WO2019/204766). Additional repressor domains that can be appended to a Cas protein are provided in, for example, WO2022/140577; Alerasool, et al., *Nature Methods*, 1093-1096, 2020.

For examples of some of the above polypeptides used in the context of fusions with Cas9, Zinc Finger, and/or TALE proteins (for site specific target nucleic modification, modulation of transcription, and/or target protein modification, e.g., histone modification), see, e.g.: Nomura *et al.*, *J Am Chem Soc.* 2007 , 129(28):8676-7; Rivenbark *et al.*, *Epigenetics.* 2012,7(4):350-60; *Nucleic Acids Res.* 2016, *44*(12):5615-28; Gilbert *et al.*, *Cell.* 2013, *154*(2):442-51; Kearns et al., *Nat Methods.* 2015 *12*(5):401-3; Mendenhall *et al.*, *Nat Biotechnol.* 2013, *31*(12):1133-6; Hilton et. al., *Nat Biotechnol.* 2015, *33*(5):510-7; Gordley et. al., *Proc Natl Acad Sci USA.* 2009, *106*(13):5053-8; Akopian *et. al.*, *Proc Natl Acad Sci USA.* 2003, *100*(15):8688-91; Tan *et al.*, *J Virol.* 2006, *80*(4):1939-48; Tan *et al.*, *Proc Natl Acad Sci USA.* 2003;*100*(21):11997-2002; Papworth *et al.*, *Proc Natl Acad Sci USA.* 2003, *100*(4):1621-6; Sanjana *et al.*, *Nat Protoc.* 2012, *7*(1):171-92; Beerli *et al.*, *Proc Natl Acad Sci USA.* 1998, *95*(25):14628-33; Snowden *et al.*, *Curr Biol.* 2002, *12*(24):2159-66; Xu *et al.*, *Cell Discov.* 2016, 2:16009; Komor *et al.*, *Nature.* 2016;*533*(7603):420-4;

Chaikind *et al.*, *Nucleic Acids Res.* 2016, *44*(20):9758–9770; Choudhury *et al.*, *Oncotarget.* 2016, *7*:46545-46556; Du *et al.*, *Cold Spring Harb Protoc.* 2016, *2016*(1):pdb.prot090175; Pham *et al.*, *Methods Mol Biol.* 2016, *1358*:43-57; Balboa *et al.*, *Stem Cell Reports.* 2015, *5*(3):448-59; Hara *et al.*, *Sci Rep.* 2015, *5*:11221; Piatek *et al.*, *Plant Biotechnol J.* 2015, *13*(4):578-89; Hu *et al.*, *Nucleic Acids Res.* 2014, *42*(7):4375-90; Cheng *et al.*, *Cell Res.* 2013, *23*(10):1163-71; Cheng *et al.*, *Cell Res.* 2013,*23*(10):1163-71; and Maeder *et al.*, *Nat Methods.* 2013 Oct, *10*(10):977-9.

In certain embodiments, the heterologous domain may comprise a permeant domain to promote uptake by a cell. A number of permeant domains are known in the art. For example, WO2017/106569 and US20180363009A1, incorporated by reference herein in entirety, describe fusion of a Cas protein with one or more nuclear localization sequences (NLS) to facilitate cell uptake. In other embodiments, a permeant peptide may be derived from the third alpha helix of *Drosophila melanogaster* transcription factor Antennapaedia, referred to as penetratin. As another example, the permeant peptide comprises the HIV-1 tat basic region amino acid sequence, which may include, for example, amino acids 49-57 of naturally-occurring tat protein. Other permeant domains include poly-arginine motifs, for example, the region of amino acids 34-56 of HIV-1 Rev protein, nona-arginine, octa-arginine, and the like. The site at which the fusion is made may be selected in order to optimize the biological activity, secretion or binding characteristics of the polypeptide. The optimal site can be determined by routine experimentation.

In certain embodiments, the heterologous domain may comprise a Protein Transduction Domain ("PTD"), also known as a cell penetrating peptide ("CPP"). A PTD is a polypeptide that facilitates traversing a lipid bilayer, micelle, cell membrane, organelle membrane, or vesicle membrane. In certain embodiments, a PTD fused to a Cas protein facilitates the Cas protein traversing a membrane, for example going from an extracellular space to an intracellular space, or from the cytosol to within an organelle. In some embodiments, a PTD is covalently linked to the amino terminus of a Cas protein. In some embodiments, a PTD is covalently linked to the carboxyl terminus of a Cas protein. In some embodiments, the PTD is inserted internally in the sequence of a Cas protein at a suitable insertion site. In some embodiments, a Cas protein includes (is conjugated to, is fused to) one or more PTDs (*e.g.*, two or more, three or more, four or more PTDs). In some embodiments, a PTD includes one or more nuclear localization signals (NLS). Examples of PTDs include but are not limited to peptide transduction domain of HIV TAT, a polyarginine sequence comprising a number of arginines sufficient to direct entry into a cell (*e.g.*, 3, 4, 5, 6, 7, 8, 9, 10, or 10-50 arginines); a VP22 domain (Zender et al. (2002) Cancer Gene Ther. 9(6):489-96), a *Drosophila Antennapedia* protein transduction domain (Noguchi et al. (2003) Diabetes 52(7): 1732-1737), a truncated human calcitonin peptide (Trehin et al. (2004) Pharm. Research 21 :1248-1256) or polylysine (Wender et al. (2000) Proc. Natl. Acad. Sci. USA 97: 13003-13008). In some embodiments, the PTD is an activatable CPP (ACPP) (Aguilera et al. (2009) Integr Biol (Camb) June; 1(5-6): 371-381). ACPPs comprise a polycationic CPP (*e.g.*, Arg9 or "R9") connected via a cleavable linker to a matching polyanion (*e.g.*, Glu9 or "E9"), which reduces the net charge to nearly zero and thereby inhibits adhesion and uptake into cells. Upon cleavage of the linker, the polyanion

is released, locally unmasking the polyarginine and its inherent adhesiveness, thus "activating" the ACPP to traverse the membrane.

In certain embodiments, a Cas protein can be linked to a heterologous polypeptide (a heterologous amino acid sequence) via a linker polypeptide (*e.g.*, one or more linker polypeptides). In some embodiments, a Cas protein can be linked at the C-terminal and/or N-terminal end to a heterologous polypeptide (fusion partner) via a linker polypeptide (*e.g.*, one or more linker polypeptides). The linker polypeptide may have any of a variety of amino acid sequences. Proteins can be joined by a spacer peptide, generally of a flexible nature, although other chemical linkages are not excluded. Suitable linkers include polypeptides of between 4 amino acids and 40 amino acids in length, or between 4 amino acids and 25 amino acids in length. 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. The ordinarily skilled artisan will recognize that design of a peptide conjugated to any elements described above 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.

2. Certain Subcellular Localization Sequences

In certain embodiments, a guided nucleic acid binding agent or a Cas fusion protein may comprise one or more subcellular localization sequences, such as nuclear localization sequences. A nuclear localization sequence is a short sequence of amino acids that promotes uptake of the protein into the nucleus of a cell. A variety of eukaryotic nuclear localization signal (NLS) sequences have been used to improve the nuclear uptake of proteins (see, e.g., Lu, et al. *Cell Comm. and Signalling*, 2021). One or more NLS may be located at the N-terminus and/or the C-terminus of a polypeptide, or may be located at any point within the polypeptide sequence. In some embodiments, the one or more NLS located at the N-terminus are identical to the one or more NLS located at the C-terminus. In some embodiments, the one or more NLS located at the N-terminus are different to the one or more NLS located at the C-terminus. In some embodiments, the one or more NLS can be located within 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acids to the N- or C-terminus. In some embodiments, the one or more NLS can be linked to the N- or C-terminus by a linker peptide. In some embodiments, an NLS is linked to another NLS by a linker. In other embodiments, the NLS linked to the N-terminus are different to the NLS linked to the C-terminus.

A monopartite classical NLS has the sequence K (K/R) X (K/R), wherein K is lysine and X can be any amino acid. A bipartite NLS comprises two clusters of 2-3 positively charged amino acids separated by a region of 9-12 amino acids that comprises several prolines, with a consensus sequence of $R/K(X)_{10-12}KRXK$, wherein R is arginine, K is lysine, and X is any amino acid. NLS (or multiple NLSs) are of sufficient strength to drive accumulation of an engineered Cas protein 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 a Cas 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.

In certain embodiments, a Cas protein may comprise one or more nuclear localization sequences selected from the table below or a variant thereof. Other NLS are known in the art (see, e.g., WO 2023/235818).

Table 1

Nuclear Localization Sequences

Nuclear Localization Sequence	SEQ ID NO:
PKLKRQ	470
RPRK	471
RRARRPRG	472
GKRKLITSEEERSPAKRGRKS	473
KGKKGRTQKEKKAARARSKGKN	474
RKRCAAGVGGGPAGCPAPGSTPLKKPRR	475
RKPVTAQERQREREEKRRRRQERAKEREKRRQERER	476
RSGGNHRRNGRGGRGGYNRRNNGYHPY	477
TLLLRETMNNLGVSDHAVLSRKTPQPY	478
PGKMDKGEHRQERRDRPY	479
GKKKKGKPGKRREQRKKKRRT	480
SANKVTKNKSNSSPYLNKRKGKPGPDS	481
VHSHKKKKIPTSPTFTTPKTLTLRRQPKYPRKSAPRRNKLDHY	482
RKHKTNRKPR	483
NRRAKAKR	484
RNKKKK	485
KRPAATKKAGQAKKKK	486
PAAKRVKLD	487

In certain embodiments, a Cas protein may deliberately not include one or more NLS so that the protein is not targeted to the nucleus, which can be advantageous; *e.g.*, when the target nucleic acid is an RNA that is present in the cytosol.

In certain embodiments, a Cas protein may include one or more subcellular localization sequences, such as nuclear export sequences. A nuclear export sequence promotes exclusion of the protein from the nucleus of a cell.

In certain embodiments, a Cas protein may include one or more subcellular localization sequences, such as a sequence to keep the fusion protein retained in the cytoplasm, a mitochondrial localization signal for targeting to the mitochondria, a chloroplast localization signal for targeting to a chloroplast, or an ER retention signal.

In certain embodiments, a Cas protein may include a tag (e.g., the heterologous domain is a detectable label) for ease of tracking and/or purification (*e.g.*, a fluorescent protein, *e.g.*, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), mCherry, tdTomato, and the like; a histidine tag, *e.g.*, a 6XHis tag; a hemagglutinin (HA) tag; a FLAG tag; a Myc tag; and the like).

3. Engineered Cas Variants with Altered PAM Specificity

A guided nucleic acid binding agent that comprises a Cas protein binds to target DNA at a sequence defined by the region of complementarity between the target-recognition region, or spacer, of the guide and the target DNA. Site-specific binding (and/or cleavage) of a double stranded target DNA for Cas proteins typically occurs at a site determined by both: (i) hybridization of the target-recognition region (spacer) of the guide to the target DNA, and (ii) the presence of a short motif (referred to as the PAM, or protospacer adjacent motif) within the non-complementary strand of DNA. In certain embodiments, the PAM for a Cas protein is immediately 5' of the protospacer sequence. In certain embodiments, the PAM for a Cas protein is immediately 3' of the protospacer sequence.

Methods for engineering the PAM interacting domain (PI domain) for Cas proteins, including SpCas9, SaCas9, StlCas9, FnCas9, AsCas12a, among others, have been previously described (see, e.g., WO2016/141224; Table 1 in Liu, *et al.*, *Trends Biotechnol. 2021 39*(3):262-273).

A bacterial selection system can be used to optimize a Cas protein variant with a non-native PI domain configured to recognize a variant PAM. Bacteria are transfected with a plasmid encoding an inducible toxic gene as well as a plasmid encoding a Cas protein variant. After inducing the toxic gene, the only surviving bacteria are those in which the Cas protein variant cleaves the toxic gene. Using this method, a large library of Cas proteins with randomized mutations within the PI domain can be screened against a large randomized PAM library. For surviving clones, a bacterial-based site-depletion assay can then be used to profile the PAM specificities of the evolved Cas proteins. These methods have been previously used to identify PI domains with novel PAM recognition sequences for SpCas9, SaCas9, StlCas9 (Kleinstiver, *Nature*, 523(7561): 481-485, 2015); SaCas9 (Kleinstiver, *Nature Biotech.*, 33(12):1293-1298, 2015). Similarly, a selection system using phage-assisted continuous evolution (PACE) has been used to identify

Cas9 variants with expanded PAM recognition (Hu, et al., *Nature,* 556: 57-63, 2918; Miller, et al., *Nature Biotech.,* 38: 471-481, 2020).

Another method that has been previously used to engineer alternative PAM interacting domains that recognize a variant PAM is structure-guided protein engineering. Using this method, amino acids that form contacts with the target DNA are identified via crystal structures, and Cas protein variants containing mutations of these amino acids are then tested for activity. A variant of saCas9 with decreased off-target editing was identified using these methods (Tan, et al., *PNAS,* 116(42):20969-20976, 2019), as was a variant of AsCas12a (Kleinstiver, et al., *Nature Biotech.,* 37(3): 276-282, 2019). This method was also used to identify a variant of FnCas9 with a broader PAM recognition sequence (Hirano, et al., *Cell*, 164:950-961).

A further method that has been used to generate Cas protein variants that recognize a variant PAM is to generate a domain-swapped chimera using direct substitution of the PI domain of a Cas protein with the PI domain of a closely related orthologous Cas protein. Early work showed that swapping the PI domain of SpCas9 with the PI domain of St3Cas9 to generate a variant of St3Cas9 that cleaved target DNA with the canonical SpCas9 5'-NGG-3' PAM (Nishimasu, et al, Cell, 156(5):935-949, 2014). In a similar manner, several Cas9 variants that recognize variant PAMs were identified by swapping PI domains with closely related Cas proteins (Ma, et al., *Nature Comm.*, 10:560, 2019).

C. Certain Nucleic Acids Encoding Polypeptides

1. Certain Nucleic Acids

In certain embodiments, provided herein are nucleic acids encoding polypeptides. In certain embodiments, the polypeptide is or is part of a guided nucleic acid binding agent. In certain embodiments, the polypeptide comprises or consists of a Cas protein.

In certain embodiments, the nucleic acid is a DNA. In certain embodiments, the nucleic acid is an RNA. In certain embodiments, the nucleic acid is an exogenous mRNA.

In certain embodiments, an exogenous mRNA comprises a coding region (e.g., an open reading frame (ORF)) encoding a polypeptide sequence, a cap, and one or more non-coding regions. In certain embodiments, the exogenous mRNA comprises a cap, 5' UTR, 3' UTR, a coding region, a poly(A) tail, and optionally one or more introns. The mRNA may comprise nucleotides selected from adenosine, guanosine, cytosine, uridine, N1-methylpseudouridine, optionally selected from adenosine, guanosine, cytosine, and uridine.

Naturally-occurring eukaryotic mRNA molecules can contain non-coding regions, including, but not limited to untranslated regions (UTR) at their 5 '-end (5' UTR) and/or at their 3 '-end (3' UTR), a 5'-cap structure and a 3'-poly(A) tail. Exogenous mRNA may be configured to include such regions, which facilitate cellular processing and translation. In certain embodiments, a formulation, e.g., an LNP may be configured to

deliver an exogenous mRNA having an open reading frame encoding a polypeptide. In certain embodiments, provided are methods in which an exogenous mRNA is translated in a cell.

An exogenous mRNA may comprise a repeat region of contiguous nucleosides such as adenosine nucleosides, e.g., a poly(A) tail. A poly(A) tail is a region that is 3' of the 3' UTR that contains multiple, consecutive adenosine monophosphates. It is believed that a poly(A) tail may protect mRNA from enzymatic degradation, e.g., in the cytoplasm, and aids in transcription termination, and/or export of the mRNA from the nucleus, and in translation. A poly(A) tail may contain 10 to 300 contiguous adenosine nucleosides. For example, an exogenous mRNA may comprise at its 3' terminus a repeat region of 50-100, 50-150, 50-200, 100-250, 120-160, or 200-300 adenosine monophosphates.

2. Capping Groups

Eukaryotic mRNAs, as well as RNA viral genomes, include a 7-methylguanosine (m7G) cap at the 5′ end of the mRNA sequence, attached to the 5'-most mRNA nucleotide through a 5′,5′-triphosphate bridge (ppp) during mRNA in vitro transcription. The cap structure plays essential functions in mRNA translation by recruiting translation initiation factors, and different 5′ caps can be incorporated into naturally occurring mRNAs. Cap0 protects endogenous mRNA from nuclease attack and is also involved in nuclear export and translation initiation. Both Cap1 and Cap2 are two 5′ caps that contain additional methyl groups on the second or third ribonucleotide. The additional modification of Cap1 and Cap2 is believed to reduce immunogenicity compared to Cap0.

An exogenous mRNA may comprise a specific capping group such as described herein or as known in the art. Examples of 5' capping groups include: "Cap 0": m7G(5')ppp(5')N; "Cap 1": m7G(5')ppp(5')(2'OMeN); and "Cap 2": m7G(5')ppp(5')(2'OMeN)(2'OMeN); in which m7G indicates a guanosine nucleoside methylated at its 7-position and having a free 3'-OH, (5') indicates a 5' point of attachment, p is a phosphate linkage, each N is independently a nucleoside, e.g. guanosine or adenosine, and (2'OMeN) is independently a 2-O-methyl nucleoside, e.g., 2'-O-methylguanosine or 2'-O-methyladenosine. A first nucleoside adenosine following the cap may also be methylated at its N6 position.

Specific capping groups include (m7(3′OMeG)(5′)ppp(5′)(2′OMeA)pG; 3'-O-Me-m7G(5')ppp(5')G ("ARCA" cap); G(5')ppp(5')A; G(5')ppp(5')G; m7G(5')ppp(5')A; and m7G(5')ppp(5')(2'OMeN)pG ("CleanCap™"), where m7(3′OMeG) indicates a guanosine nucleoside methylated at its 7-position and having a 3'-O-methyl. Commercial sources are available for some groups (e.g., New England BioLabs, Ipswich, MA, and TriLink Biotechnologies, San Diego, CA). 5'- Capping of exogenous mRNA may be completed post-transcriptionally, e.g., using Vaccinia Virus Capping Enzyme. Cap 1 structure may be generated using both Vaccinia Virus Capping Enzyme and a 2'-O methyl-transferase to generate. Cap 2 structure may be generated from the Cap 1 structure followed by the 2'-O-methylation of the 5'- third most nucleotide using a 2'-O methyl-transferase. Cap 3 structure may be generated from the Cap 2 structure followed by the 2'-O-methylation of the 5'-fourthmost nucleotide using a 2'-O methyl-transferase. Enzymes may be derived, for example, from a recombinant source.

3. Poly(A) Tail

The 3'-poly(A) tail is a region of contiguous adenine nucleotides at the 3 '-end of the transcribed mRNA. In certain embodiments, the 3'-poly(A) tail comprises one to 400 adenine nucleotides. In certain embodiments, the 3'-poly(A) tail comprises unmodified adenine nucleosides linked by phosphodiester internucleoside linkages.

In some embodiments, the exogenous mRNA comprises a stabilizing element. A stabilizing element may be a histone stem-loop. A stem-loop binding protein (SLBP), a 32 kDa protein has been identified. It is associated with the histone stem- loop at the 3 '-end of the histone messages in both the nucleus and the cytoplasm, and is believed to promote efficient 3 '-end processing of histone pre-mRNA and stimulation of translation. In some embodiments, the histone stem-loop sequence comprises a length of 15 to 45 nucleotides.

In some embodiments, exogenous mRNA includes a coding region, at least one histone stem-loop, and optionally, a poly(A) sequence or polyadenylation signal. The poly(A) sequence or polyadenylation signal is believed to increase the expression of an encoded protein. The encoded protein, in some embodiments, is not a histone protein, a reporter protein (e.g. Luciferase, GFP, EGFP, b-Galactosidase, EGFP), or a marker or selection protein (e.g. alpha-Globin, Galactokinase and Xanthine:guanine phosphoribosyl transferase (GPT)).

4. Coding Region and Sequence Optimization

A coding region may comprise a contiguous sequence beginning with a start codon (e.g., methionine (AUG)), and ending with a stop codon (e.g., UAA, UAG or UGA), and comprising one or more nucleotide sequences encoding amino acids. Generally, a coding region encodes a polypeptide which forms a protein *in vivo* upon administration to a subject.

In certain embodiments, a coding region comprises selected nucleotide codons to improve one or more properties of the exogenous mRNA or of the encoded protein, and may comprise a codon optimized ORF. For example, a coding region of an exogenous mRNA sequence may be codon optimized. Codon optimization, in some embodiments, may be used to match codon frequencies in target and host organisms to ensure proper folding; bias GC content to increase mRNA stability or reduce secondary structures; minimize tandem repeat codons or base runs that may impair gene construction or expression; customize transcriptional and translational control regions; insert or remove protein trafficking sequences; add or remove post translation modification sites in encoded protein (e.g., glycosylation sites); add, remove or shuffle protein domains; insert or delete restriction sites; modify ribosome binding sites and mRNA degradation sites; adjust translational rates to allow the various domains of the protein to fold properly; or generally to reduce or eliminate problem secondary structures within the polynucleotide. Codon optimization tools, algorithms and services are known in the art - non limiting examples include services from GeneArt (Life Technologies), DNA2.0 (Menlo Park CA) and/or proprietary methods. In certain embodiments, a coding region sequence is optimized using an optimization algorithm as provided herein or as known in the art.

In certain embodiments, a coding region encodes a polypeptide. In certain embodiments, the encoded polypeptide is or is part of a guided nucleic acid binding agent. In certain embodiments, the encoded polypeptide comprises or consists of a guided nucleic acid binding agent. In certain embodiments, the guided nucleic acid binding agent comprises a Cas protein.

In certain embodiments, the encoded polypeptide is a variant that differs in amino acid sequence from a wild-type (naturally occurring), native, or reference protein sequence, and may possess substitutions, deletions, and/or insertions at certain positions within the amino acid sequence, as compared to a wild-type, native or reference sequence. In certain embodiments, the % identity to a wild-type, native or reference sequence is in a range of X to Y, wherein X represents the lowest number in the range and Y represents the highest number in the range, wherein X and Y are each independently selected from 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100; provided that $X \leq Y$.

In certain embodiments, variant polypeptides encoded by nucleic acids may contain amino acid changes that confer any one or more desirable properties, such as reducing clearance, enhancing immunogenicity, enhancing expression, and/or improving stability or pharmacokinetic/pharmacodynamic ("PK/PD") properties. Variant polypeptides can be made using routine mutagenesis techniques and assayed to determine the presence of desired properties. In addition, assays to determine expression levels and immunogenicity are well known, and PK/PD properties of a variant polypeptide can also be measured using well known techniques, such as by determining protein expression in a subject over time. Stability of a variant polypeptide can be measured, for example, by assaying thermal stability, stability upon urea denaturation or using *in silico* prediction.

In certain embodiments, variant polypeptides encoded by nucleic acids may contain sequence tags or other additional amino acids at the N-terminal or C-terminal ends. These can be used for peptide detection, purification or localization. For example, lysines can be used to increase peptide solubility or to allow for biotinylation. Alternatively, amino acids located at the carboxy and amino terminal regions of the amino acid sequence of a polypeptide may be deleted, thereby providing for truncated sequences (fragments). In some embodiments, cavities in the core of proteins can be filled to improve stability, *e.g.*, by introducing larger amino acids. In other embodiments, buried hydrogen bond networks may be replaced with hydrophobic resides to improve stability. In yet other embodiments, glycosylation sites may be removed and replaced with appropriate alternative amino acids. Such sequences are readily identifiable to one of skill in the art.

In certain embodiments, the encoded polypeptide is a fusion protein. The fusion protein may include two or more proteins or fragments thereof joined together. In certain embodiments, the encoded polypeptide is a Cas fusion protein.

In certain embodiments, a coding region further encodes a linker located between at least one or each domain of a fusion protein. In some embodiments, the linker is a cleavable linker or protease-sensitive linker. In some embodiments, the linker is selected from the group consisting of F2A linker, P2A linker, T2A linker, E2A linker, and combinations thereof. This family of self-cleaving peptide linkers, referred to as 2A peptides, has been described in the art (see for example, Kim, J.H. et al. (2011) *PLoS ONE* 6:e18556 and WO2017/127750). In some embodiments, the linker is an F2A linker. In some embodiments, the linker is a GGGS linker.

In certain embodiments, a coding region comprises a signal peptide fused to a polypeptide of interest. Signal peptides, comprising the N-terminal 15-60 amino acids of proteins, are typically required for translocation across the membrane on the secretory pathway and, thus, universally control the entry of most proteins to the secretory pathway. The signal peptide of a nascent precursor protein (pre-protein) directs the ribosome to the rough endoplasmic reticulum (ER) membrane and initiates the transport of the growing peptide chain across it for processing. ER processing produces mature proteins, wherein the signal peptide is cleaved from precursor proteins, typically by an ER-resident signal peptidase of the host cell, or they remain uncleaved and function as a membrane anchor. A signal peptide may also facilitate targeting of a polypeptide to the cell membrane. Signal peptides from heterologous genes are known in the art and can be tested for desired properties and then incorporated into a nucleic acid.

In certain embodiments, a signal peptide may have a length of X-Y amino acids, wherein X represents the lowest number in the range and Y represents the highest number in the range, wherein X and Y are each independently selected from 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60; provided that $X \leq Y$.

In certain embodiments, a coding region shares less than 95%, less than 90%, less than 80%, or less than 70% sequence identity to a naturally-occurring or wild-type sequence ORF (e.g., a naturally-occurring or wild-type mRNA sequence encoding a protein), or a range of values therebetween.

In some embodiments, an exogenous mRNA may be one in which the levels of G/C are enhanced. The G/C-content of nucleic acid molecules (e.g., mRNA) may influence the stability of the RNA. RNA having an increased amount of guanine (G) and/or cytosine (C) nucleotides may be more stable than RNA containing a large amount of adenine (A) and thymine (T) or uracil (U) nucleotides. See, e.g., WO02/098443.

The coding region comprises at least one start codon (an AUG nucleotide triplet) among its 5'-most nucleotides and at least one stop codon (an UAG, UAA, or UGA nucleotide triplet) among its 3'-most nucleotides. In certain embodiments, an exogenous mRNA includes 2 or more stop codons, e.g., 2-10 stop codons.

Exemplary codons for specific amino acids are known in the art.

5. Chemical Modifications For Exogenous mRNA

The compositions of the present disclosure comprise, in some embodiments, an exogenous mRNA comprising modified nucleotides or nucleosides (those other than A, C, G, and U joined by only phosphodiester linkages; exclusive of 5' cap). Such modified nucleotides and nucleosides can be naturally-occurring modified nucleotides and nucleosides or non-naturally occurring modified nucleotides and nucleosides. Such modifications can include those described herein with respect to oligonucleotides.

In certain embodiments, a naturally-occurring modified nucleotide or nucleoside of the disclosure is one generally known or recognized in the art. Non-limiting examples of such naturally occurring modified nucleotides and nucleosides can be found, *inter alia*, in the widely recognized MODOMICS database. In certain embodiments, an exogenous mRNA comprises a modified nucleotide or modified nucleoside described in one or more of US application Nos. PCT/US2012/058519; PCT/US2013/075177; PCT/US2014/058897; PCT/US2014/058891; PCT/US2014/070413; PCT/US2015/027400; PCT/US2015/36773; PCT/US2015/36759; PCT/US2015/36771; or PCT/IB 2017/051367 all of which are incorporated by reference herein. Hence, an exogenous mRNA can comprise unmodified nucleosides, modified nucleosides, or a combination thereof. In certain embodiments, each nucleoside in an exogenous mRNA is modified similarly to other nucleosides of the same type; for example, all uridine nucleosides in a parent sequence are replaced by N1-methylpseudouridine nucleosides. A modification may provide reduced degradation and/or reduced immunogenicity compared to an RNA comprising only unmodified nucleosides.

In certain embodiments, an exogenous mRNA comprises a modified nucleoside selected from 1-methylpseudouridine, 1-ethylpseudouridine, 5-methoxyuridine, 5-methylcytidine, and pseudouridine. In certain embodiments, an exogenous mRNA comprises a modified nucleoside selected from 5-methoxymethyl uridine, 5-methylthiouridine, 1-methoxymethyl pseudouridine, 5-methylcytidine, and 5-methoxycytidine. In certain embodiments, an exogenous mRNA comprises a modified nucleoside selected from a combination of two or more (e.g., 2, 3, 4) of any of the modified nucleobases of this paragraph.

In certain embodiments, an exogenous mRNA comprises 1-methylpseudouridine in place of one or more, e.g. all uridine nucleosides of a parent sequence. In certain embodiments, an exogenous mRNA comprises 1-methylpseudouridine at one or more, e.g., all uridine positions of a parent sequence, and 5-methylcytidine at one or more, e.g., all cytidine positions of a parent sequence.

In certain embodiments, an exogenous mRNA comprises pseudouridine at one or more, e.g. all uridine nucleosides of a parent sequence. In certain embodiments, an exogenous mRNA comprises pseudouridine at one or more, e.g., all uridine positions of a parent sequence, and 5-methylcytidine at one or more, e.g., all cytidine positions of a parent sequence.

In certain embodiments, an exogenous mRNA comprises unmodified uridine at one or more, e.g., all uridine positions of the nucleic acid.

In certain embodiments, all nucleotides of a particular type in an exogenous mRNA (or in a sequence region thereof) are modified nucleotides compared to a parent sequence, wherein the nucleotides modified

from the parent sequence may be any one of nucleotides A, G, U, C, or any one of the combinations A+G, A+U, A+C, G+U, G+C, U+C, A+G+U, A+G+C, G+U+C or A+G+C.

In certain embodiments, a nucleic acid may contain from X%-Y% modified nucleotides (either in relation to overall nucleotide content, or in relation to one or more types of nucleotide, *i.e.*, any one or more of A, G, U or C), wherein X represents the lowest number in the range and Y represents the highest number in the range, wherein X and Y are each independently selected from 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100; provided that $X \leq Y$.

6. Untranslated Regions (UTRs)

In wild-type mRNA, certain regions of a nucleic acid may be transcribed into RNA but not translated. In exogenous mRNA as described herein, a 5' UTR may start at the transcription start site and continues to a start codon at the beginning of the coding region, but does not include the start codon. The 3' UTR following a stop codon may include a transcription termination signal. The 5' UTR and the 3' UTR do not encode protein (are non-coding regions).

In certain embodiments, the presence of a UTR enhances the stability of the exogenous mRNA, or induces downregulation of the exogenous mRNA in undesirable sites or tissues. A variety of 5' UTR and 3' UTR sequences are known and available in the art.

A 5' UTR is a region of an mRNA that is directly upstream (5') from the start codon (the first codon of an mRNA transcript translated by a ribosome). In certain embodiments, the 5' UTR may provide translation initiation, and/or form secondary structures which are involved in elongation factor binding. The 5' UTR may include an initiation sequence such as a Kozak sequence. The Kozak sequence, also called the Kozak consensus sequence, is believed to be involved in ribosomal initiation of translation. In certain embodiments, the Kozak sequence comprises AUGG. In certain embodiments, the Kozak sequence is GCCRCCAUGG (SEQ ID NO: 468) or CCRCCAUGG, where R is a purine nucleoside (adenosine or guanosine).

In certain embodiments of the disclosure, a 5' UTR is an unmodified UTR, i.e., one found in nature. In another embodiment, a 5' UTR is a modified UTR, i.e., does not occur in nature. In certain embodiments, a modified UTR increases gene expression relative to an unmodified counterpart. Exemplary 5' UTRs include *Xenopus laevis* or human derived a-globin or b-globin (e.g., US Patent No. 9012219), human cytochrome b-245 a polypeptide, and hydroxysteroid (17b) dehydrogenase, and Tobacco etch virus (e.g., US9012219), CMV immediate-early 1 (IE1) gene (US2014/0206753, WO2013/185069), the sequence GGGAUCCUACC (SEQ ID NO: 469) (WO2014/144196), and a 5' UTR described in US Patent Application Publication No. 2010/0293625 or WO2015085318. In another embodiment, 5' UTR of a TOP gene is a 5' UTR of a TOP gene

lacking the 5' TOP motif (the oligopyrimidine tract) (e.g., WO/2015/101414, W02015/101415, WO/2015/062738, WO2015/024667, WO2015/024667); 5' UTR element derived from ribosomal protein Large 32 (L32) gene (WO/2015/101414, W02015/101415, WO/2015/062738), 5' UTR element derived from the 5' UTR of an hydroxysteroid (17-b) dehydrogenase 4 gene (HSD17B4) (WO2015/024667), or a 5' UTR element derived from the 5' UTR of ATP5A1 (WO2015/024667) can be used. In some embodiments, an internal ribosome entry site (IRES) is used instead of a 5' UTR.

In general, an exogenous mRNA may include a UTR from any suitable gene. The naturally occurring UTR or portions thereof may be placed in the same orientation as in the transcript from which they were selected or may be altered in orientation or location. Hence a 5' or 3' UTR may be inverted, shortened, or lengthened. A reference UTR, e.g., a naturally occurring UTR, may be altered by the inclusion of additional nucleotides, deletion of nucleotides, swapping, or transposition of nucleotides or sequences thereof.

In certain embodiments, the 3' UTR sequences may include adenosine- and uridine-containing repeats. Such AU rich sequences are believed to provide high rates of turnover. Based on their sequence features and functional properties, the AU rich elements (AREs) can be separated into three classes (Chen et al, 1995): Class I AREs contain several dispersed copies of an AUUUA motif within U-rich regions. C-Myc and MyoD contain class I AREs. Class II AREs possess two or more overlapping UUAUUUASS nonamers, where S is adenosine or uridine. Molecules containing this type of AREs include GM-CSF and TNF-a. Class III ARES are less well defined. These U rich regions do not contain an AUUUA motif. c-Jun and Myogenin are two well-studied examples of this class. It is believed that including the HuR specific binding sites into the 3' UTR will provide stabilization of the message *in vivo*. In certain embodiments, a 3' UTR includes a repeated ARE.

Unmodified (natural) and modified (non-natural) 3' UTR sequences are known in the art. UTRs known in the art include globin UTRs, including Xenopus b-globin UTRs and human b-globin UTRs (9012219, US2011/0086907), a modified b-globin construct (US2012/0195936, WO2014/071963), a2-globin, al-globin, UTRs (WO2015/101415, WO2015/024667), CYBA (Ferizi *et al.*, 2015) and albumin (Thess *et al.*, 2015), bovine or human growth hormone (wild type or modified) (WO2013/185069, US2014/0206753, WO2014152774), rabbit b globin and hepatitis B virus (HBV), a-globin 3' UTR and Viral VEEV 3' UTR sequences, the sequence UUUGAAUU (WO2014/144196), human and mouse ribosomal protein, rps93'UTR (WO2015/101414), FIG4 (WO2015/101415), and human albumin 7 (WO2015/101415).

The untranslated region may also include an upregulation motif, e.g., a translation enhancer element (TEE). As a non-limiting example, the TEE may include those described in WO1999024595, WO2012009644, WO2009075886, WO2007025008, WO1999024595, European Patent Publication No. EP2610341A1 and EP2610340A1, US Patent No. US6310197, US6849405, US7456273, US7183395, US Patent Publication No. US20090226470, US20110124100, US20070048776, US20090093049, or US20130177581 each of which is herein incorporated by reference in its entirety.

In certain embodiments, the 5' UTR or 3' UTR may comprise one or more additional functional regions selected from an upregulation region or a ribosome binding region. In certain embodiments, the functional region is an upregulation region (e.g., TEE). In certain embodiments, the TEE is one known in the art, e.g., in US Application No. 2009/0226470. In certain embodiments, an exogenous mRNA comprises a plurality of upregulation motifs which may be the same or different from each other and which number, e.g., 2, 3, 4, 5, or more.

In certain embodiments, the exogenous mRNA comprises a ribosome binding region, e.g., an internal ribosome entry site (IRES). The IRES may be, for example, one described in US Patent No. US7468275 and International Patent Publication No. WO2001/055369, each of which is herein incorporated by reference in its entirety. In certain embodiments, the IRES is one known in the art, e.g., in WO2014/081507.

In certain embodiments, an exogenous mRNA may comprise a double, triple or quadruple UTR such as a 5' UTR or 3' UTR. As used herein, a "double" UTR is one in which two copies of the same UTR are included contiguously or substantially contiguously. For example, an exogenous mRNA may comprise a double beta-globin 3' UTR as described in US Patent publication No. 2010/0129877, which is incorporated herein by reference in its entirety.

In certain embodiments, a 5' UTR or 3' UTR may comprise a repeated set of functional sequences, e.g., as AA, ABAB or AABB or ABCABC or variants thereof, in which each letter, A, B, and C represent a different functional region. The pattern may be repeated once, twice, or 3 or more times.

Generally, any 5' UTR sequence and any 3' UTR sequence may be combined in a particular exogenous mRNA. Those of ordinary skill in the art will understand that 5'UTRs that are heterologous or synthetic may be used with any desired 3' UTR sequence. For example, a heterologous 5'UTR may be used with a synthetic 3'UTR or with a heterologous 3' UTR.

Other non-coding sequences may also be used as regions or subregions within an exogenous mRNA. For example, introns or portions of introns sequences may be incorporated into regions of nucleic acid of the disclosure. In certain embodiments, inclusion of intronic sequences may increase protein production as well as nucleic acid levels.

Combinations of features may be included in flanking regions and may be contained within other features.

7. Length

In certain embodiments, an exogenous mRNA includes 200 to 3,000 nucleotides. For example, an exogenous mRNA may include 200 to 500, 200 to 1000, 200 to 1500, 200 to 3000, 500 to 1000, 500 to 1500, 500 to 2000, 500 to 3000, 1000 to 1500, 1000 to 2000, 1000 to 3000, 1500 to 3000, or 2000 to 3000 nucleotides. Unless otherwise indicated, the exogenous mRNA comprises contiguous nucleosides.

8. Preparation

In some embodiments, the exogenous mRNA is generated using a non-amplified, linearized DNA template in an *in vitro* transcription reaction to generate the exogenous mRNA. In some embodiments, the

template DNA is isolated DNA. In some embodiments, the template DNA is cDNA. In some embodiments, the cDNA is formed by reverse transcription of an RNA polynucleotide.

An *in vitro* transcription system typically comprises a transcription buffer, nucleotide triphosphates (NTPs), an RNase inhibitor and a polymerase. The NTPs may be purchased from a supplier or may be synthesized as described herein according to methods known in the art.

In vitro transcription of RNA is known in the art and is described in International Publication WO 2014/152027, which is incorporated by reference herein in its entirety. In certain embodiments, the exogenous mRNA is prepared in accordance with any one or more of the methods described in WO 2015/164674, WO 2018/053209, WO 2019/036682 and WO 2022/221440, each of which is incorporated by reference herein.

In some embodiments, the exogenous mRNA is generated using solid phase chemical synthesis, liquid phase chemical synthesis, a combination of synthetic methods or via ligation. The use of solid phase or liquid phase chemical synthesis in combination with enzymatic ligation may provide efficient generation of long chain exogenous mRNA that is difficult to obtain by chemical synthesis alone.

Solid phase chemical synthesis can be an automated method wherein molecules are immobilized on a solid support and synthesized step by step in a reactant solution. Solid-phase synthesis is useful in site-specific introduction of chemical modifications.

Liquid phase chemical synthesis can also be an automated method involving sequential addition of monomer building in a liquid phase.

DNA or RNA ligases can also be used to promote intermolecular ligation of the 5' and 3' ends of polynucleotide chains through the formation of a phosphodiester bond. Nucleic acids such as chimeric polynucleotides and/or circular nucleic acids may be prepared by ligation of one or more regions or subregions. DNA fragments can be joined by a ligase catalyzed reaction to create recombinant DNA with different functions. Two oligodeoxynucleotides, one with a 5' phosphoryl group and another with a free 3' hydroxyl group, serve as substrates for a DNA ligase. The DNA can then be used as a template in an *in vitro* transcription reaction to generate the exogenous mRNA.

9. Purification

Purification of the exogenous mRNA described herein may comprise steps including nucleic acid clean-up, quality assurance, and quality control. Clean-up may be performed by methods known in the arts such as, but not limited to, AGENCOURT® beads (Beckman Coulter Genomics, Danvers, MA), poly-T beads, LNATM oligo-T capture probes (EXIQON® Inc, Vedbaek, Denmark) or HPLC based purification methods such as, but not limited to, strong anion exchange HPLC, weak anion exchange HPLC, reverse phase HPLC (RP-HPLC), and hydrophobic interaction HPLC (HIC-HPLC). The term "purified" when used in relation to exogenous mRNA refers to one that is separated from at least one contaminant. A

"contaminant" is any substance that makes another unfit, impure or inferior. Thus, a purified nucleic acid (e.g., DNA and RNA) is present in a form or setting different from that in which it is found in nature, or a form or setting different from that which existed prior to subjecting it to a treatment or purification method.

A quality assurance and/or quality control check may be conducted using methods such as, but not limited to, gel electrophoresis, UV absorbance, or analytical HPLC.

In some embodiments, the nucleic acids may be sequenced by methods including, but not limited to reverse-transcriptase-PCR. In certain embodiments, the exogenous mRNA comprises an open reading frame that encodes a polypeptide sequence represented by any one of SEQ ID NO.: 4-95.

II. Certain Oligomeric Agents

In certain embodiments, provided herein are oligonucleotides, which consist of linked nucleosides. In certain embodiments, a single guide comprises an oligonucleotide. Oligonucleotides may be unmodified oligonucleotides or may be modified oligonucleotides. Modified oligonucleotides comprise at least one modification relative to unmodified nucleic acids. That is, modified oligonucleotides comprise at least one modified nucleoside (comprising a modified sugar moiety and/or a modified nucleobase) and/or at least one modified internucleoside linkage.

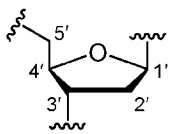
A. Certain Modified Nucleosides

Modified nucleosides comprise a modified sugar moiety or a modified nucleobase or both a modifed sugar moiety and a modified nucleobase.

1. Certain Sugar Moieties

In certain embodiments, modified sugar moieties are non-bicyclic modified sugar moieties. In certain embodiments, modified sugar moieties are bicyclic or tricyclic sugar moieties. In certain embodiments, modified sugar moieties are sugar surrogates. Such sugar surrogates may comprise one or more substitutions corresponding to those of other types of modified sugar moieties.

In certain embodiments, modified sugar moieties are non-bicyclic modified furanosyl sugar moieties comprising one or more substituent groups including, but not limited to, substituents at the 2′, 3′, 4′, and/or 5′ positions, as numbered below:



In certain embodiments, the modified furanosyl sugar moiety is a ribosyl sugar moiety that is not an unmodified sugar moiety (*i.e.*, an unmodified RNA or unmodified DNA moiety). In certain embodiments, the modified furanosyl sugar moiety is a xylosyl, lyxosyl, or arabinosyl sugar moiety.

In certain embodiments, non-bicyclic modified sugar moieties are 2′-substituted sugar moieties and comprise a substituent group at the 2′-position. In certain embodiments one or more non-bridging substituent of non-bicyclic modified sugar moieties is branched. Examples of substituent groups suitable for the 2′-position of modified sugar moieties include but are not limited to: 2′-F, 2′-OCH_3 ("OMe" or "O-methyl"), and 2′-$O(CH_2)_2OCH_3$ ("MOE" or "O-methoxyethyl"). In certain embodiments, 2′-substituent groups are selected from: halo, allyl, amino, azido, SH, CN, OCN, CF_3, OCF_3, C_1-C_{10} alkoxy, substituted C_1-C_{10} alkoxy, C_1-C_{10} alkyl, substituted C_1-C_{10} alkyl, S-alkyl, N(R_m)-alkyl, O-alkenyl, S-alkenyl, N(R_m)-alkenyl, O-alkynyl, S-alkynyl, N(R_m)-alkynyl, O-alkylcnyl-O-alkyl, alkynyl, alkaryl, aralkyl, O-alkaryl, O-aralkyl, $O(CH_2)_2SCH_3$, $O(CH_2)_2ON(R_m)(R_n)$ or $OCH_2C(=O)$-N(R_m)(R_n), where each R_m and R_n is, independently, H, an amino protecting group, or substituted or unsubstituted C_1-C_{10} alkyl, $O(CH_2)_2ON(CH_3)_2$ ("DMAOE"), or 2′-$O(CH_2)_2O(CH_2)_2N(CH_3)_2$ ("DMAEOE"). Synthetic methods for some of these 2′-substituent groups may be found, *e.g.*, in Cook *et al.*, U.S. 6,531,584; Cook *et al.*, U.S. 5,859,221; and Cook *et al.*, U.S. 6,005,087. Certain embodiments of these 2′-substituent groups may be further substituted with one or more substituent groups independently selected from: halo, cyano, OR^{a2}, NO_2, NH_2, NHR^{a2}, $N(R^{a2})_2$, C_1-C_6 alkyl, C_1-C_6 haloalkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_{10} cycloalkyl, C_6-C_{10} aryl, heteroaryl, heterocyclyl, C_1-C_6 alkylene-NH_2, C_1-C_6alkylene-NHR^{a2}, C_1-C_6 alkylene-$N(R^{a2})_2$, $C(O)R^{a3}$, $C(O)OR^{a3}$, $C(O)NHR^{a3}$, $C(O)N(C_1$-C_4 alkyl)R^{a3}, SR^{a3}, $S(O)_2R^{a3}$, $S(O)R^{a3}$, $NHC(O)R^{a3}$, $N(C_1$-C_4 alkyl)$C(O)R^{a3}$, $NHS(O)R^{a3}$, $N(C_1$-C_4alkyl)$S(O)R^{a3}$, $NHS(O)_2R^{a3}$, and $N(C_1$-C_4 alkyl)$S(O)_2R^{a3}$; each R^{a2} is independently selected from C_2-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_{10} cycloalkyl, C_6-C_{10} aryl, heteroaryl, and heterocyclyl; each R^{a3} is independently hydrogen, OH, C_1-C_6 alkyl, C_1-C_6 haloalkyl, C_3-C_{10} cycloalkyl, C_6-C_{10} aryl, heteroaryl, or heterocyclyl. In certain embodiments, a sugar moiety comprises two of the above substituents at the 2′-position. In certain embodiments, a sugar moiety comprises a 2′-fluoro and a second 2′-substituent.

In certain embodiments, a 2′-substituted sugar moiety comprises a non-bridging 2′-substituent group selected from: F, NH_2, N_3, OCF_3, OCH_3, $O(CH_2)_3NH_2$, $CH_2CH{=}CH_2$, $OCH_2CH{=}CH_2$, $OCH_2CH_2OCH_3$, $O(CH_2)_2SCH_3$, $O(CH_2)_2ON(R_m)(R_n)$, $O(CH_2)_2O(CH_2)_2N(CH_3)_2$, and N-substituted acetamide ($OCH_2C(=O)$-N(R_m)(R_n)), where each R_m and R_n is, independently, H, an amino protecting group, or substituted or unsubstituted C_1-C_{10} alkyl. In certain embodiments, a 2′-substituted sugar moiety comprises a non-bridging 2′-substituent group selected from: F, OCF_3, OCH_3, $OCH_2CH_2OCH_3$, $O(CH_2)_2SCH_3$, $O(CH_2)_2ON(CH_3)_2$, $O(CH_2)_2O(CH_2)_2N(CH_3)_2$, $O(CH_2)_2ON(CH_3)_2$ ("DMAOE"), $O(CH_2)_2O(CH_2)_2N(CH_3)_2$ ("DMAEOE"), and $OCH_2C(=O)$-N(H)CH_3 ("NMA").

In certain embodiments, a 2′-substituted sugar moiety comprises a 2′-substituent group selected from: F, OCH_3, and $OCH_2CH_2OCH_3$.

In certain embodiments, modified furanosyl sugar moieties and nucleosides incorporating such modified furanosyl sugar moieties are further defined by stereochemical configuration. For example, a 2′-deoxyfuranosyl sugar moiety (*i.e.*, 2′-(H)H furanosyl sugar moiety) may be in seven isomeric configurations

other than the naturally occurring β-D-deoxyribosyl configuration. Such modified sugar moieties are described in, *e.g.*, WO 2020/072991, incorporated by reference herein. A 2′-modified sugar moiety has an additional stereocenter at the 2′-position relative to a 2′-deoxyfuranosyl sugar moiety; therefore, such sugar moieties have a total of sixteen possible stereochemical configurations. Modified furanosyl sugar moieties described herein are in the β-D-ribosyl stereochemical configuration unless otherwise specified.

In certain embodiments, non-bicyclic modified sugar moieties comprise a substituent group at the 4′-position. Examples of substituent groups suitable for the 4′-position of modified sugar moieties include, but are not limited to, alkoxy (*e.g.*, methoxy), alkyl, and those described in Manoharan *et al.*, WO 2015/106128.

In certain embodiments, non-bicyclic modified sugar moieties comprise a substituent group at the 3′-position. Examples of substituent groups suitable for the 3′-position of modified sugar moieties include, but are not limited to, alkoxy (*e.g.*, methoxy), alkyl (*e.g.*, methyl, ethyl).

In certain embodiments, non-bicyclic modified sugar moieties comprise a substituent group at the 5′-position. Examples of substituent groups suitable for the 5′-position of modified sugar moieties include, but are not limited to, vinyl, alkoxy (*e.g.*, methoxy), alkynyl, allyl, and alkyl (*e.g.*, methyl (*R* or *S*), ethyl (*R* or *S*)).

In certain embodiments, non-bicyclic modified sugar moieties comprise more than one non-bridging sugar substituent, for example, 2′-F-5′-methyl sugar moieties, such as described in Migawa *et al.*, US 2010/0190837, which is incorporated herein by reference, or alternative 2′- and 5′-modified sugar moieties as described in Rajeev *et al.*, US 2013/0203836.

Certain modified sugar moieties are bicyclic sugar moieties and comprise a substituent that bridges two atoms of the furanosyl ring to form a second ring. In certain embodiments, the bicyclic sugar moiety comprises a bridge between the 4′ and the 2′ furanose ring atoms. Examples of such 4′ to 2′ bridging sugar substituents include, but are not limited to: 4′-CH_2-2′, 4′-$(CH_2)_2$-2′, 4′-$(CH_2)_3$-2′, 4′-CH_2-O-2′ (“LNA”), 4′-CH_2-S-2′, 4′-$(CH_2)_2$-O-2′ (“ENA”), 4′-$CH(CH_3)$-O-2′ (referred to as “constrained ethyl” or “cEt” when in the *S* configuration), 4′-CH_2-O-CH_2-2′, 4′-CH_2-N(R)-2′, 4′-$CH(CH_2OCH_3)$-O-2′ (“constrained MOE” or “cMOE”) and analogs thereof, 4′-$C(CH_3)(CH_3)$-O-2′ and analogs thereof, 4′-CH_2-$N(OCH_3)$-2′ and analogs thereof, 4′-CH_2-O-$N(CH_3)$-2′, 4′-CH_2-$C(H)(CH_3)$-2′, 4′-CH_2-$C(=CH_2)$-2′ and analogs thereof, 4′-$C(R_aR_b)$-N(R)-O-2′, 4′-$C(R_aR_b)$-O-N(R)-2′, 4′-CH_2-O-N(R)-2′, and 4′-CH_2-N(R)-O-2′, wherein each R, R_a, and R_b is, independently, H, a protecting group, or C_1-C_{12} alkyl. Representative U.S. patents that teach the preparation of such bicyclic sugar moieties include, but are not limited to: Imanishi *et al.*, U.S. 7,427,672; Swayze *et al.*, U.S. 7,741,457; Swayze *et al.*, U.S. 8,022,193; Seth *et al.*, U.S. 8,278,283; Prakash *et al.*, U.S. 8,278,425; and Seth *et al.*, U.S. 8,278,426.

2.Certain Modified Nucleobases

In certain embodiments, modified oligonucleotides comprise one or more nucleoside comprising an unmodified nucleobase. In certain embodiments, modified oligonucleotides comprise one or more nucleoside comprising a modified nucleobase. In certain embodiments, modified oligonucleotides comprise one or more nucleoside that does not comprise a nucleobase, referred to as an abasic nucleoside. In certain embodiments, modified oligonucleotides comprise one or more inosine nucleosides (i.e., nucleosides comprising a hypoxanthine nucleobase). An "unmodified nucleobase" is adenine (A), thymine (T), cytosine (C), uracil (U), or guanine (G). A modified nucleobase is a group of atoms other than unmodified A, T, C, U, or G capable of pairing with at least one other nucleobase. A 5-methylcytosine is an example of a modified nucleobase. A universal base is a modified nucleobase that can pair with any one of the five unmodified nucleobases.

In certain embodiments, modified nucleobases are selected from: 5-substituted pyrimidines, 6-azapyrimidines, alkyl or alkynyl substituted pyrimidines, alkyl substituted purines, and N-2, N-6 and O-6 substituted purines. In certain embodiments, modified nucleobases are selected from: 5-methylcytosine, 2-aminopropyladenine, 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-N-methylguanine, 6-N-methyladenine, 2-propyladenine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-propynyl (-C≡C-CH_3) uracil, 5-propynylcytosine, 6-azouracil, 6-azocytosine, 6-azothymine, 5-ribosyluracil (pseudouracil), N1-methylpseudouracil, 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl, 8-aza and other 8-substituted purines, 5-halo (particularly 5-bromo), 5-trifluoromethyl, 5-halouracil, and 5-halocytosine, 7-methylguanine, 7-methyladenine, 2-F-adenine, 2-aminoadenine, 7-deazaguanine, 7-deazaadenine, 3-deazaguanine, 3-deazaadenine, 6-N-benzoyladenine, 2-N-isobutyrylguanine, 4-N-benzoylcytosine, 4-N-benzoyluracil, 5-methyl 4-N-benzoylcytosine, 5-methyl 4-N-benzoyluracil, universal bases, hydrophobic bases, promiscuous bases, size-expanded bases, and fluorinated bases. Further modified nucleobases include tricyclic pyrimidines, such as 1,3-diazaphenoxazine-2-one, 1,3-diazaphenothiazine-2-one and 9-(2-aminoethoxy)-1,3-diazaphenoxazine-2-one (G-clamp). Modified nucleobases may also include those in which the purine or pyrimidine base is replaced with other heterocycles, for example 7-deaza-adenine, 7-deazaguanosine, 2-aminopyridine and 2-pyridone. Further nucleobases include those disclosed in Englisch *et al.*, *Angew Chem Int Ed.* 1991, 30, 613; Sanghvi, Y.S., Chapter 15, *Antisense Research and Applications*, Crooke, S.T. and Lebleu, B., *Eds.*, CRC Press, 1993, 273-288; and those disclosed in Chapters 6 and 15, *Antisense Drug Technology*, Crooke S.T., Ed., CRC Press, 2008, 163-166 and 442-443.

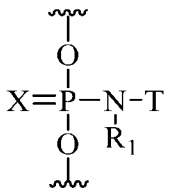
Publications that teach the preparation of certain of the above noted modified nucleobases, as well as other modified nucleobases include without limitation, Rogers et al., U.S. 5,134,066 ; Benner et al., U.S. 5,432,272; Matteucci et al., U.S. 5,502,177 ; Froehler et al., U.S. 5,594,121 ; and Cook et al., U.S. 5,681,941.

B. Certain Modified Internucleoside Linkages

The naturally occurring internucleoside linkage of RNA and DNA is a 3' to 5' phosphodiester linkage. In certain embodiments, nucleosides of modified oligonucleotides may be linked together using one

or more modified internucleoside linkages. The two main classes of internucleoside linking groups are defined by the presence or absence of a phosphorus atom. Representative phosphorus-containing internucleoside linkages include, but are not limited to, phosphodiesters, which contain a phosphodiester bond (also referred to as unmodified or naturally occurring linkages), phosphotriesters, methylphosphonates, phosphoramidates, phosphorothioates, phosphonoacetates ("PACE"), thiophosphonoacetates ("Thio-PACE") and phosphorodithioates. Representative non-phosphorus containing internucleoside linking groups include, but are not limited to, methylenemethylimino (-CH_2-N(CH_3)-O-CH_2-), thiodiester, thionocarbamate (-O-C(=O)(NH)-S-); siloxane (-O-SiH_2-O-); and N,N'-dimethylhydrazine (-CH_2-N(CH_3)-N(CH_3)-). Modified internucleoside linkages, compared to naturally occurring phosphodiester internucleoside linkages, can be used to alter, typically increase, nuclease resistance of the oligonucleotide. In certain embodiments, internucleoside linkages having a chiral atom can be prepared as a racemic mixture, or as separate enantiomers. Methods of preparation of phosphorous-containing and non-phosphorous-containing internucleoside linkages are well known to those skilled in the art.

In certain embodiments, a modified internucleoside linkage is any of those described in WO2021/030778, incorporated by reference herein. In certain embodiments, a modified internucleoside linkage comprises the formula:



wherein independently for each such internucleoside linking group of a modified oligonucleotide:

X is selected from O or S;

R_1 is selected from H, C_1-C_6 alkyl, and substituted C_1-C_6 alkyl; and

T is selected from SO_2R_2, C(=O)R_3, and P(=O)R_4R_5, wherein:

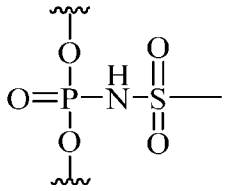
R_2 is selected from an aryl, a substituted aryl, a heterocycle, a substituted heterocycle, an aromatic heterocycle, a substituted aromatic heterocycle, a diazole, a substituted diazole, a C_1-C_6 alkoxy, C_1-C_6 alkyl, C_1-C_6 alkenyl, C_1-C_6 alkynyl, substituted C_1-C_6 alkyl, substituted C_1-C_6 alkenyl substituted C_1-C_6 alkynyl, and a linker;

R_3 is selected from an aryl, a substituted aryl, CH_3, N(CH_3)$_2$, OCH_3 and a linker;

R_4 is selected from OCH_3, OH, C_1-C_6 alkyl, substituted C_1-C_6 alkyl and a linker; and

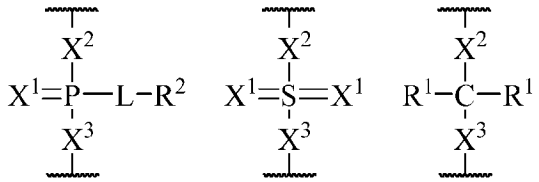
R_5 is selected from OCH_3, OH, C_1-C_6 alkyl, and substituted C_1-C_6 alkyl.

In certain embodiments, a modified internucleoside linkage comprises a mesyl phosphoramidate linking group having a formula:



In certain embodiments, a modified internucleoside linkage is a neutral internucleoside linkage. Neutral internucleoside linkages include, without limitation, phosphotriesters, methylphosphonates, MMI (3'-CH_2-N(CH_3)-O-5'), amide-3 (3'-CH_2-C(=O)-N(H)-5'), amide-4 (3'-CH_2-N(H)-C(=O)-5'), formacetal (3'-O-CH_2-O-5'), methoxypropyl (MOP), and thioformacetal (3'-S-CH_2-O-5'). Further neutral internucleoside linkages include nonionic linkages comprising siloxane (dialkylsiloxane), carboxylate ester, carboxamide, sulfide, sulfonate ester and amides (See for example: *Carbohydrate Modifications in Antisense Research*; Y.S. Sanghvi and P.D. Cook, Eds., ACS Symposium Series 580; Chapters 3 and 4, 40-65). Further neutral internucleoside linkages include nonionic linkages comprising mixed N, O, S and CH_2 component parts.

In certain embodiments, a modified internucleoside linkage is a standard length internucleoside linkage, and is represented by Formula Z1, Z2, or Z3 below. As used herein, a "standard length internucleoside linkage" refers to an internucleoside linkage that has a structure represented by Formula Z1, Z2, or Z3:



Z1 Z2 Z3

wherein independently for each internucleoside linking group of Formula Z1, Z2, or Z3:

each X^1 is independently selected from O and S;

X^2 is selected from O, NR^1, CH_2, and S;

X^3 is selected from O, NR^1, CH_2, and S;

L is absent, NR^1, $N(R^1)SO_2$, –N=, O, C_1-C_6 alkylene, or C_1-C_6 heteroalkylene;

each R^1 is independently selected from H, C_1-C_6 alkyl, and substituted C_1-C_6 alkyl, or two R^1 on the same atom together form =O; and

R^2 is selected from -OH, -SH, C_1-C_{22} alkyl, substituted C_1-C_{22} alkyl, C_2-C_{22} alkenyl, substituted C_2-C_{22} alkenyl, cycloalkyl, substituted cycloalkyl, heterocyclyl, substituted heterocyclyl, heteroaryl, substituted heteroaryl, aryl, and substituted aryl;

wherein when a group is substituted, it comprises one or more substituent groups selected from halo, -OH, -$N(R^1)_2$, -O-C_1-C_6 alkyl, C_1-C_{22} alkyl, C_2-C_{22} alkenyl, cycloalkyl, heterocyclyl, heteroaryl, and aryl.

In certain embodiments, an extended internucleoside linkage is an extended length internucleoside linkage. For example, in certain embodiments, an extended internucleoside linkage forms when two functional groups on two oligonucleotides react to link the two oligonucleotides together to form a single oligonucleotide comprising an extended internucleoside linkage. One such reaction is the click reaction between bicyclo[6.1.0]nonyne and an azide. Additional linkers suitable for joining two oligonucleotides with Click chemistry are described in "Click Chemistry for Biotechnology and Materials Science" Ed. Joerg Lahann, Wiley 2009. Further examples of linking chemistry include an inverse electron demand Diels-Alder reaction, *e.g.*, as described in Argamunt *et al.*, *J. Org. Chem.* 2020, *85*, 10, 6593–6604, Sarrett *et al.*, *Nat. Protocols* 2021, *16*, 3348–3381; Handula *et al.*, *Molecules*, 2021, *26* (15), 4640, Wiessler *et al.*, *Int. J. Med. Sci.* 2010, *7* (1), 19–28; copper-catalyzed azide-alkyne cycloaddition (CuAAC) see, e.g., S. I. Presolski, *et al.*, *J. Am. Chem. Soc.* 2010, 132, 14570–14576; D. Soriano Del Amo, *et al.*, *J. Am. Chem. Soc.*, 2010, *132*, 16893–16899; Staudinger reaction, see, *e.g.*, Saxon and C. R. Bertozzi, *Science*, 2000, *287*, 2007–2010; B. L. Nilsson, *et al.*, *Org. Lett.*, 2000, *2*, 1939–1941, E. Saxon, *et al.*, *Org. Lett.*, 2000, *2*, 2141–2143; formation of hydrazones and oximes, see, *e.g.*, J. Y. Axup, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, *109*, 16101–16106; photoclick reactions, see, *e.g.*, W. Song, *et al.*, *Angew. Chem. Int. Ed.*, 2008, *47*, 2832–2835, A. Herner and Q. Lin, *Top. Curr. Chem.*, 2016, *374*, 1; strain-promoted alkyne-nitrone cycloaddition (SPANC) reactions, see, *e.g.*, D. A. MacKenzie, *et al.*, *Curr. Opin. Chem. Biol.*, 2014, *21*, 81–88; transition metal catalyzed cross coupling, see, *e.g.*, M. Chalker, *et al.*, *J. Am. Chem. Soc.*, 2009, *131*, 16346–16347; nucleophilic additions, in particular, of a thiol to a maleimide, see, *e.g.*, Kang *et al.*, *Chem. Sci.*, 2021, *12*, 13613-13647, Bernardim *et al.*, *Nat. Comm.* 2016, *7*, 13128, Jain *et al.*, *Pharm. Res.* 2015, *32* (11), 3526-3540.

C. Certain Motifs

In certain embodiments, guides (modified oligonucleotides) comprise one or more modified nucleoside(s) comprising a modified sugar moiety. In certain embodiments, modified oligonucleotides comprise one or more modified nucleosides comprising a modified nucleobase. In certain embodiments, modified oligonucleotides comprise one or more modified internucleoside linkage. In such embodiments, the modified, unmodified, and differently modified sugar moieties, nucleobases, and/or internucleoside linkages of a modified oligonucleotide define a pattern or motif. In certain embodiments, the patterns of sugar moieties, nucleobases, and internucleoside linkages are each independent of one another. Thus, a modified oligonucleotide may be described by its sugar motif, nucleobase motif and/or internucleoside linkage motif

(as used herein, nucleobase motif describes the modifications to the nucleobases independent of the sequence of nucleobases).

1. Certain Sugar Motifs

In certain embodiments, oligonucleotides comprise one or more type of modified sugar and/or unmodified sugar moiety arranged along the oligonucleotide or region thereof in a defined pattern or sugar motif. In certain instances, such sugar motifs include but are not limited to any of the sugar modifications discussed herein. Generally, the guide comprises unmodified RNA nucleosides, and optionally one or more modified nucleosides as described herein. In certain embodiments, the guide comprises or consists of a modified oligonucleotide.

In certain embodiments, the 5'-most nucleoside of a modified oligonucleotide comprises a modified sugar moiety. In certain embodiments, the 3'-most nucleoside of a modified oligonucleotide comprises a modified sugar moiety. In certain embodiments, at least 1, 2, 3, 4, or 5 of the 5'-most five nucleosides of a modified oligonucleotide comprise a modified sugar moiety. In certain embodiments, at least 1, 2, 3, 4, or 5 of the 3'-most five nucleosides of a modified oligonucleotide comprise a modified sugar moiety.

In certain embodiments, the above modifications (sugar, nucleobase, internucleoside linkage) are incorporated into a modified oligonucleotide. In certain embodiments, modified oligonucleotides are characterized by their modification motifs and overall lengths as described herein, and such parameters are each independent of one another. Unless otherwise indicated, all modifications are independent of nucleobase sequence. In certain embodiments, a modified oligonucleotide is a guide. In certain embodiments, the guide has a sugar moiety, nucleobase, and/or internucleoside linkage motif described in any of the following references, each of which are hereby incorporated by reference: WO 2014/144761, WO 2015/026885, WO 2016/089433, WO 2016/100951, WO 2016/123230, WO 2016/164356, WO 2017/004261, WO 2017/004279, WO 2017/068377, WO 2017/136794, WO 2017/181107, WO 2017/214460, WO 2018/009822, WO 2018/057946, WO 2018/098383, WO 2018/107028, WO 2018/125964, WO 2019/084664, WO 2019/147275, WO 2019/147743, WO 2019/183000, WO 2019/237069, WO 2021/119006, WO 2021/119275, WO 2021/125840, WO 2021/207651 , WO 2021/207711, WO 2022/086846.

2. Certain Nucleobase Motifs

In certain embodiments, oligonucleotides comprise modified and/or unmodified nucleobases arranged along the oligonucleotide or region thereof in a defined pattern or motif. In certain embodiments, each nucleobase is modified. In certain embodiments, none of the nucleobases are modified. In certain embodiments, each purine or each pyrimidine is modified. In certain embodiments, each adenine is modified. In certain embodiments, each guanine is modified. In certain embodiments, each thymine is modified. In certain embodiments, each uracil is modified. In certain embodiments, each cytosine is modified. In certain embodiments, some or all of the cytosine nucleobases in a modified oligonucleotide are 5-methyl cytosines.

In certain embodiments, all of the cytosine nucleobases are 5-methyl cytosines and all of the other nucleobases of the modified oligonucleotide are unmodified nucleobases.

In certain embodiments, modified oligonucleotides comprise a block of modified nucleobases. In certain such embodiments, the block is at the 3'-end of the oligonucleotide. In certain embodiments the block is within 3 nucleosides of the 3'-end of the oligonucleotide. In certain embodiments, the block is at the 5'-end of the oligonucleotide. In certain embodiments the block is within 3 nucleosides of the 5'-end of the oligonucleotide.

In certain embodiments, the modified nucleobase is selected from pseudouracil, 2-6 diamino purine, 2-thiouracil, 4-thiouracil, 2-aminoadenine, 6-methyladenine, hypoxanthine, N-1-methylpseudouracil, or 5-methylcytosine.

3. Certain Internucleoside Linkage Motifs

In certain embodiments, oligonucleotides comprise modified and/or unmodified internucleoside linkages arranged along the oligonucleotide or region thereof in a defined pattern or motif. In certain embodiments, each internucleoside linking group is a phosphodiester internucleoside linkage (P=O). In certain embodiments, each internucleoside linking group of a modified oligonucleotide is a phosphorothioate internucleoside linkage (P=S). In certain embodiments, each internucleoside linkage of a modified oligonucleotide is independently selected from a phosphorothioate internucleoside linkage and phosphodiester internucleoside linkage. In certain embodiments, each internucleoside linkage of a modified oligonucleotide is independently selected from a phosphorothioate internucleoside linkage, phosphodiester internucleoside linkage, and a phosphonoacetate internucleoside linkage.

In certain embodiments, the 5'-most internucleoside linkage of a modified oligonucleotide is a modified internucleoside linkage. In certain embodiments, the 3'-most internucleoside linkage of the guide is a modified internucleoside linkage. In certain embodiments, at least 1, 2, 3, 4, or 5 of the 5'-most five internucleoside linkages of a modified oligonucleotide are modified internucleoside linkages. In certain embodiments, at least 1, 2, 3, 4, or 5 of the 3'-most five internucleoside linkages of a modified oligonucleotide are modified internucleoside linkages. In certain embodiments, the modified internucleoside linkage is a phosphorothioate internucleoside linkage.

4. Certain Guide Chemical Modification Motifs

A guide can comprise any of the modifications commonly applied to modified oligonucleotides described herein, including modified sugar moieties, modified internucleoside linkages, and modified nucleobases. In certain embodiments, a guide is modified at the 3'-end, the 5'-end, or both the 3'-end and 5'-end. In certain embodiments, the three nucleosides at the 3'-end and the three nucleosides at the 5'-end are

2'-OMe nucleosides, and the remainder of the nucleosides of the guide are unmodified RNA nucleosides. In certain embodiments, each nucleoside within the 5'-most stem loop is an unmodified RNA nucleoside.

In certain embodiments, each internucleoside linkage of the guide is an unmodified phosphodiester linkage. In certain embodiments, the guide comprises one or more modified internucleoside linkages. In certain embodiments, the guide comprises one or more phosphorothioate internucleoside linkages. In certain embodiments, the 3'-most one, two, three, four, or five internucleoside linkages are modified internucleoside linkages. In certain embodiments, the 5'-most one, two, three, four, or five internucleoside linkages are modified internucleoside linkages. In certain embodiments, the 3'-most and 5'-most one, two, three, four, or five internucleoside linkages are modified internucleoside linkages, and the remainder of the linkages are unmodified phosphodiester internucleoside linkages. In certain embodiments, the 3'-most one, two, three, four, or five internucleoside linkages are phosphorothioate internucleoside linkages. In certain embodiments, the 5'-most one, two, three, four, or five internucleoside linkages are phosphorothioate internucleoside linkages. In certain embodiments, the 3'-most and 5'-most one, two, three, four, or five internucleoside linkages are phosphorothioate internucleoside linkages, and the remainder of the linkages are unmodified phosphodiester internucleoside linkages. In certain embodiments, the three 3'-most and the three 5'-most internucleoside linkages are phosphorothioate internucleoside linkages, and the remainder of the linkages are unmodified phosphodiester internucleoside linkages.

D. Certain Lengths

In certain embodiments, oligonucleotides (including guides) can have any of a variety of ranges of lengths. In certain embodiments, oligonucleotides consist of X to Y linked nucleosides, where X represents the fewest number of nucleosides in the range and Y represents the largest number nucleosides in the range. In certain such embodiments, X and Y are each independently selected from 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180 ; provided that $X \leq Y$. For example, in certain embodiments, oligonucleotides consist of 50-180, 80-180, 90-180, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 160-170, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150, 100-140, 110-140, 110-130, 120-140, or 120-150 linked nucleosides.

In certain embodiments, a single guide oligonucleotide consists of 50-180, 80-180, 90-180, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 160-170, 90-100,100-110, 110-120, 120-130, 130-140, 140-150, 100-140, 110-140, 110-130, 120-140, or 120-150 linked nucleosides. In certain embodiments,

the target-recognition region or spacer of a guide consists of 15-30, 18-26, 18-24, 18-22, 20-26, 20-24, 20-22, 21-23, 22-23, 20, 21, 22, or 23 linked nucleosides.

In certain embodiments, the protein-recognition region of a guide or scaffold consists of 60-130, 60-120, 60-110, 60-100, 70-130, 70-120, 70-110, 70-100, 70-90, 70-80, 80-130, 80-120, 80-110, 80-100, 80-90, 90-130, 90-120, 90-110, 90-100, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, or 110 linked nucleosides. In certain embodiments, the protein-recognition region of a guide or scaffold represents a truncated version of a native guide for the corresponding Cas protein. In certain embodiments, the guide is truncated at the 3' end. In certain embodiments, the guide is truncated at the 5' end. In certain embodiments, the guide is shortened by removing nucleotides from linker, stem loop and/or hairpin structures, such as removing one base pair from a hairpin to shorten the hairpin. In certain embodiments, the truncated version is at least 5, at least 10, at least 15, or at least 20 nucleotides shorter than the native guide. In certain embodiments, the protein-recognition region or scaffold of a guide represents a length-extended version of a native guide for the corresponding Cas protein. In certain embodiments, additional nucleotides are added at the 3' end. In certain embodiments, additional nucleotides are added at the 5' end. In certain embodiments, additional nucleotides are inserted within linkers, stem loops and/or hairpins of the native guide.

E. Nucleobase Sequence

In certain embodiments, oligonucleotides (unmodified or modified oligonucleotides) are further described by their nucleobase sequence. In certain embodiments oligonucleotides have a nucleobase sequence that is complementary to a second oligonucleotide or an identified reference nucleic acid, such as a target nucleic acid. In certain such embodiments, a region of an oligonucleotide has a nucleobase sequence that is complementary to a second oligonucleotide or an identified reference nucleic acid, such as a target nucleic acid. In certain embodiments, the nucleobase sequence of a region, such as a target-recognition region, or entire length of an oligonucleotide is at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95%, or 100% complementary to the second oligonucleotide or nucleic acid, such as a target nucleic acid.

F. Secondary Structure

In certain embodiments, oligonucleotides or portions thereof adopt a defined secondary structure. In certain embodiments, secondary structure is determined by Watson-Crick base pairing, Hoogsteen base pairing, and/or non-canonical base pairing interactions. The secondary structure of an oligonucleotide sequence can be predicted using standard software, such as the Vienna RNA package RNAfold (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi; see Lorenz, et al., *Algorithms for Molecular Biology*, 6:1 26, 2011). In certain embodiments, an oligonucleotide may have more than one predicted secondary structure.

A guide oligonucleotide has secondary structure, at least some portion of which is important in forming contacts with its cognate Cas protein.

A "spacer region" of a guide, also known as a "target recognition region", has a spacer sequence. In certain embodiments, the spacer sequence is a sequence is complementary to the target sequence of a target nucleic acid. In some embodiments, the spacer sequence ranges from X to Y nucleobases, wherein X represents the lowest number in the range and Y represents the highest number in the range, wherein X and Y are each independently selected from 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,47, 48, 49 or 50; provided that X≤Y. In some embodiments, a spacer sequence contains 18, 19 or 20 nucleobases. In some embodiments, the guide hybridizes, or is capable of hybridizing, to the forward strand of the target nucleic acid. In some embodiments, the guide hybridizes, or is capable of hybridizing, to the reverse strand of the target nucleic acid.

In certain embodiments, a target sequence is in a region of the target nucleic acid that is on the opposite strand of a DNA from a PAM sequence. The spacer hybridizes to the complementary region located in the non-PAM strand of the target nucleic acid. The spacer region interacts with a target nucleic acid of interest in a sequence-specific manner via hybridization. The nucleotide sequence of the spacer thus varies depending on the target sequence of the target nucleic acid.

In certain embodiments, the spacer region is designed to hybridize to a region of the target nucleic acid that is complementary to a region located 5' of a PAM recognizable by guided nucleic acid binding agent. In certain embodiments, the guided nucleic acid binding protein is a Cas protein. The spacer sequence can perfectly match the target sequence or can have mismatches. Each guided nucleic acid binding agent has a particular PAM sequence that it recognizes in a target nucleic acid, though some guided nucleic acid binding agents can recognize PAM sequences with variation at one or more positions.

In some embodiments, the target sequence ranges from X to Y nucleotides, wherein X represents the lowest number in the range and Y represents the highest number in the range, wherein X and Y are each independently selected from 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,47, 48, 49 or 50; provided that X≤Y. In some embodiments, a target sequence contains 18, 19 or 20 nucleotides.

In some embodiments, the percent complementarity between the spacer sequence and the target nucleic acid sequence can be about, at least, at least about, at most or at most about 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%. In some embodiments, the spacer sequence and the target sequence is 100% complementary. In some embodiments, the percent complementarity between the spacer sequence and the sequence is 100% over the six contiguous 5'-most nucleotides of the target sequence. In some embodiments, the percent complementarity between the spacer sequence and the target nucleic acid is at least 60% over about 20 contiguous nucleotides. In other

embodiments, the spacer sequence and the target sequence can contain up to 10 mismatches, *e.g.*, up to 9, up to 8, up to 7, up to 6, up to 5, up to 4, up to 3, up to 2, or up to 1 mismatch.

The guide secondary structure features for binding to Cpf1 have been described for several analogs (see, e.g. Zetsche, *et al.*, *Cell*. 2015, *163*, 759–771). Features include a 5' direct repeat region that is the core of the protein-recognition region (or 'scaffold') of the guide followed by a 3' target recognition region. The direct repeat region forms a hairpin. In certain embodiments, a native type V Cas guide includes additional nucleotides to the 5' of the direct repeat region. In certain embodiments, a guide includes engineered nucleotides to the 5' of the direct repeat region. In certain embodiments, a guide includes engineered nucleotides to the 3' of the target recognition region. In certain such embodiments, these engineered nucleotides are a "stabilizing region". In certain such embodiments, these engineered nucleotides are a "5' stabilizing region". In certain such embodiments, these engineered nucleotides are a "3' stabilizing region".

In certain embodiments, a guide comprises a target-recognition region, a direct repeat region, and a stabilizing region appended to the 5'- or 3'-end either directly or by a linker sequence. In certain embodiments, the linker sequence is 1, 2, 3, 4, or 5 uridine nucleotides.

In certain embodiments, a "5' stabilizing region" has a sequence selected from the table below. In certain embodiments, a "3' stabilizing region" has a sequence selected from the table below.

Table 2

Guide Stabilizing Sequences

Stabilizing Sequence ID	Sequence	Reference	SEQ ID NO.
dR6	GGUGUCUGUGCCUAUGGCACAGACACCAAA ACCAAAA	Zhang *et al.*, *Nat. Chem. Biol.* 2021, *17*: 828–836.	488
dR19	GUGCACAGCUCAGUGCACAAAACCAAAA	Zhang *et al.*, *Nat. Chem. Biol.* 2021, *17*: 828–836.	489
dtR19	GUGCACAGCUCAGUGCAC	Zhang *et al.*, *Nat. Chem. Biol.* 2021, *17*: 828–836.	498
C9SL5	GAGACUUGCGAUUGCAAGUCUC	N/A	490
C57SL5	GCACGACUUGCUUUGCAAGUCGUGC	N/A	491
SRV-1 pseudoknot	GCGGCCAGCUCCAGGCCGCCAAACAAUAUGG AGC	Lim *et al.*, *J Phys Condens Matter*. 2015, *27*: 354101. PDB: 1E95	492
MS2	ACAUGAGGAUCACCCAUGU	Konermann *et al.*; *Nature*. 2015, *517*(7536):583-8	493
MALAT1	GGAAGGUUUUUCUUUUCCUGAGGCGAAAGU CUCAGGUUUUGCUUUUUGGCCUUUCUUAAA AAAAAAAAAAGCAAAA	Brown *et al.*, *Nat. Struct. Mol. Biol.* 2014, *21*(7): 633-40.	494
V071_dR	CCUAAAGCCGGCAUA	N/A	1172
V051_dR	GAAUAAUGAUUCCCUU	N/A	1173
V050_dR	GUUUAGCUACUCCUUA	N/A	1174
AsCas12a_dR	GUCAAAAGACCUUUU	N/A	1175

III. Certain Editing Systems and Expression Systems

In certain embodiments, an editing system comprises a guided nucleic acid binding agent and a guide. In certain embodiments, a guide consists of a single oligonucleotide and is a single guide. In certain embodiments, a guide comprises a target-recognition region and a protein-recognition region. In certain embodiments, the guide is a single guide consisting of an oligonucleotide that comprises a target-recognition region, or spacer, and a protein-recognition region, or scaffold. In certain embodiments, the target-recognition region is a region of an oligonucleotide that has a sequence complementary to an equal length portion of a target sequence within the complementary strand of a target DNA.

In certain embodiments, an editing system comprises a guided nucleic acid binding agent and a guide. In certain embodiments, the guided nucleic acid binding agent comprises a Cas protein. In certain embodiments, the guided nucleic acid binding agent consists of a Cas protein. In certain embodiments, the Cas protein is a Cas enzyme, dead Cas protein, or Cas fusion protein. In certain embodiments, the Cas protein has an amino acid sequence of any of SEQ ID NOs: 4-95, 501-533, 600-601. In certain embodiments, the Cas protein is encoded by DNA having the nucleobase sequence of any of SEQ ID NOs: 96-187, 534-566, 603-604. In certain embodiments, the guide is an oligonucleotide. In certain embodiments, the guide consists of two oligonucleotides. In certain embodiments, the guide comprises a protein-recognition region having a nucleobase sequence selected from any of SEQ ID NOs: 1061-1171. In certain embodiments, an editing system comprises a guided nucleic acid agent comprising a guided nucleic acid binding protein and a guide, wherein the protein-recognition region of the guide has a nucleobase sequence that binds to the guided nucleic acid binding protein. In certain embodiments, the guided nucleic acid binding agent comprises a Cas protein having an amino acid sequence of any of SEQ ID NOs: 4-95, 501-533, 600-601 and a guide comprising a protein-recognition region having the nucleobase sequence of any of SEQ ID NOs: 1061-1171, wherein the polypeptide SEQ ID NO. and the protein-recognition SEQ ID NO. are chosen from the same row of Table 52. In certain embodiments, the editing system comprises a Cas protein having an amino acid sequence that has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to any of SEQ ID NOs: 4-95 and a guide comprising a protein-recognition region having a nucleobase sequence that has 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% identity to any of SEQ ID NOs: 1061-1171. In certain such embodiments, the polypeptide SEQ ID NO. and the protein-recognition SEQ ID NO. are chosen from the same row of Table 52.

In certain embodiments, the guided nucleic acid binding agent has one or more mutations compared to a naturally-occurring guided nucleic acid binding agent. A guided nucleic acid binding agent may have one or more mutations that produce altered activity compared to a naturally occurring guided nucleic acid binding agent, such altered or abrogated DNA endonuclease activity without substantially diminished binding affinity to a target nucleic acid. Such modifications can allow for the sequence-specific nucleic acid targeting of a

guided nucleic acid binding agent for the purpose of transcriptional modulation (e.g., activation or repression); epigenetic modification or chromatin modification by methylation, demethylation, acetylation or deacetylation, or any other modifications of target nucleic acid binding and/or modifying proteins known in the art. In some embodiments, the guided nucleic acid binding agent has no DNA endonuclease activity.

In certain embodiments, the guided nucleic acid binding agent is a nickase that cleaves the complementary strand of a target DNA but has reduced ability to cleave the non-complementary strand of the target DNA, or that cleaves the non-complementary strand of the target DNA but has reduced ability to cleave the complementary strand of the target DNA. In some embodiments, the guided nucleic acid binding agent has a reduced ability to cleave both the complementary and the non-complementary strands of the target DNA.

In certain embodiments, an editing system comprises a guided nucleic acid binding agent and a guide, wherein the guide combines, or is capable of combining, with the guided nucleic acid binding agent to form a ribonucleoprotein ("RNP"). In certain embodiments, the guided nucleic acid binding agent comprises or consists of a Cas protein.

In certain embodiments, the RNP binds, or is capable of binding, to a target nucleic acid. Upon binding, the RNP may create a break in the target nucleic acid, such as a double strand break or a single strand break (e.g., a nick). In certain embodiments, the break may be repaired by a process of non-homologous end-joining ("NHEJ") or homology-directed repair ("HDR"). In certain embodiments, repair of a break can result in, for example, a gene knockout or a gene knock-in.

In certain embodiments, the guide combines, or is capable of combining, with a dead Cas protein to form a RNP that binds, or is capable of binding, to a target nucleic acid but does not break the target nucleic acid, wherein binding of the RNP prevents transcription or translation, thereby silencing expression of a target nucleic acid. In certain embodiments, the dead Cas protein is a fusion protein that includes a transcriptional repressor domain. Such embodiments encompass methods of CRISPR interference ("CRISPRi"). See, for example, Qi LS, Larson MH, et al. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression, *Cell* 152(5):1173-83, the contents of which are incorporated herein by reference in their entirety.

In certain embodiments, the guide combines, or is capable of combining, with a dead Cas protein to form a RNP that binds, or is capable of binding, to a target nucleic acid but does not break the target nucleic acid, wherein binding of the RNP activates transcription or translation, thereby activating expression of or over-expressing a target nucleic acid. In certain embodiments, the dead Cas protein is a fusion protein that includes a transcriptional activator. Such embodiments encompass methods of CRISPR activation ("CRISPRa"). See, for example, Polstein LR, Gersbach CA. (2015) A light inducible CRISPR-Cas9 system for control of endogenous gene activation, *Nat Chem Biol*, 11 (3):198-200; and Zalatan JG, Lee ME, et al. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds, *Cell* 15;160(1-2):339-50, the contents of which are incorporated herein by reference in their entirety.

In certain embodiments, the guide combines, or is capable of combining, with a dead Cas protein to form a RNP that binds, or is capable of binding, to a target nucleic acid but does not break the target nucleic acid, wherein binding of the RNP enables visualization of a target nucleic acid. In certain embodiments, the dead Cas protein is a fusion protein that includes a fluorescent protein. Such embodiments encompass methods of gene visualization. See, for example, Ma H, Naseri A, et al. (2015) Multicolor CRISPR labeling of chromosomal loci in human cells, *Proc Natl Acad Sci* USA. 10;112(10):3002- 7; and Ma H, Tu LC, et al. (2016) Multiplexed labelling of genomic loci with dCas9 and engineered sGRNAs using CRISPRainbow, *Nat Biotechnol* 34(5):528-30; and Carlson-Stevermer, J., Kelso, R., Kadina, A., Joshi, S., Rossi, N., Walker, J., Stoner, R., & Maures, T. (2020) CRISPRoff enables spatio-temporal control of CRISPR editing, *Nature Communications*, 11(1), 1-7, the contents of which are incorporated herein by reference in their entirety.

In certain embodiments, the guide combines, or is capable of combining, with a Cas protein to form a RNP that binds, or is capable of binding, to a target nucleic acid, wherein binding of the RNP enables introduction of point mutations into a target nucleic acid. In certain embodiments, the Cas protein is a fusion protein that includes a nucleobase deaminase. In certain embodiments, the nucleobase deaminase is a cytosine base editor that can introduce C>T or T>C transitions. In certain embodiments, the nucleobase deaminase is an adenine base editor that can introduce A>G or G>A transitions. Such embodiments encompass methods of base editing.

In certain embodiments, the guide is a prime editing guide and combines, or is capable of combining, with a Cas nickase protein to form a RNP that binds, or is capable of binding, to a target nucleic acid, wherein binding of the RNP enables introduction of point mutations into a target nucleic acid. In certain embodiments, the Cas nickase protein is a fusion protein that includes a reverse transcriptase. Such embodiments encompass methods of prime editing. See, for example, Gao P, Lyu Q, et al. (2021) Prime editing in mice reveals the essentiality of a single base in driving tissue-specific gene expression, *Genome Biol.* 22(1):83, the contents of which are incorporated herein by reference in their entirety.

In certain embodiments, provided is an expression system comprising an exogenous mRNA encoding a Cas protein having an amino acid sequence that has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to any of SEQ ID NOs: 4-95, 501-533, 600-601 and a guide comprising a protein-recognition region having a nucleobase sequence that has 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% identity to any of SEQ ID NOs: 1061-1171. In certain such embodiments, the polypeptide SEQ ID NO. and the protein-recognition region of SEQ ID NO. are chosen from the same row of Table 52.

IV. Certain Delivery Vehicles

A. Lipid Nanoparticles ("LNPs")

1. Composition

In certain embodiments, an editing system or an expression system can be delivered within an LNP. In certain embodiments, the LNP comprises an ionizable or cationic lipid, a non-cationic (e.g. neutral or zwitterionic lipid), a polymer-lipid, a sterol (e.g. cholesterol) or a derivative thereof, and optionally one or more excipients. The LNP at least partially encapsulates a cargo. The encapsulation of a cargo in the LNP may be 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 91% or more, 92% or more, 94% or more, or 95% or more. In some embodiments, the encapsulation of a cargo in the LNP is such that between 80% to 99%, between 80% to 97%, between 80% to 95%, between 85% to 95%, between 87% to 95%, between 90% to 95%, between 91% or more to 95% or less, 91% or more to 94% or less, over 91% to 95% or less, 92% to 99%, between 92% to 97%, or between 92% to 95% of the LNP in the encapsulates the cargo.

In certain embodiments, the cargo comprises a nucleic acid, e.g., an exogenous mRNA. In certain embodiments, the cargo comprises a guide, e.g., an oligomeric agent. In certain embodiments, the cargo comprises an exogenous mRNA and a guide. In certain embodiments, the cargo comprises different nucleic acids (*e.g.*, mRNA, gRNA, siRNA, saRNA, mcDNA, or plasmid DNA) separately or in combination, such as may be useful to provide an effect requiring colocalization of different nucleic acids. In certain embodiments, the LNP is provided as a suspension in an aqueous medium. In certain embodiments, a pharmaceutical composition comprises an LNP at least partially encapsulating a cargo in an aqueous medium.

In some embodiments, the LNP comprises 20-70 mol% ionizable or cationic lipid. For example, the LNP may comprise 20-50 mol%, 20-40 mol%, 20-30 mol%, 30-70 mol%, 30-60 mol%, 30-50 mol%, 30-40 mol%, 40-70 mol%, 40-60 mol%, 40-50 mol%, 50-70 mol% or 50-60 mol% ionizable or cationic lipid. In some embodiments, the LNP comprises 20 mol%, 30 mol%, 40 mol%, 50, 60 mol% or 70 mol% ionizable or cationic lipid.

In some embodiments, the LNP comprises 5-30 mol% non-cationic lipid. For example, the LNP may comprise 5-25 mol%, 5-20 mol%, 5-15 mol%, 5-10 mol%, 10-30 mol%, 10-25 mol%, 10-20 mol%, 10-15 mol%, 15-30 mol%, 15-25 mol%, 15-20 mol%, 20-30 mol%, 20-25 mol% or 25-30 mol% non-cationic lipid. In some embodiments, the LNP comprises 5 mol%, 10 mol%, 15 mol%, 20 mol%, 25 mol% or 30 mol% non-cationic lipid.

In some embodiments, the LNP comprises 25-55 mol% sterol. For example, the LNP may comprise 25-50 mol%, 25-45 mol%, 25-40 mol%, 25-35 mol%, 25-30 mol%, 30-55 mol%, 30-50 mol%, 30-45 mol%, 30-40 mol%, 30-35 mol%, 35-55 mol%, 35-50 mol%, 35-45 mol%, 35-40 mol%, 40-55 mol%, 40-50 mol%, 40-45 mol%, 45-55 mol%, 45-50 mol%, or 50-55 mol% sterol. In some embodiments, the LNP comprises 25 mol%, 30 mol%, 35 mol%, 40 mol%, 45 mol%, 50 mol%, or 55 mol% sterol.

In some embodiments, the LNP comprises 0.5-15 mol% polymer lipid. For example, the LNP may comprise 0.5-10 mol%, 0.5-5 mol%, 1-15 mol%, 1-10 mol%, 1-5 mol%, 2-15 mol%, 2-10 mol%, 2-5 mol%, 5-15 mol%, 5-10 mol%, or 10-15 mol% polymer lipid. In some embodiments, the LNP comprises 0.5 mol%, 1 mol%, 2 mol%, 3 mol%, 4 mol%, 5 mol%, 6 mol%, 7 mol%, 8 mol%, 9 mol%, 10 mol%, 11 mol%, 12 mol%, 13 mol%, 14 mol%, or 15 mol% polymer lipid.

In certain embodiments, the LNP comprises 30 to 70 mol % of an ionizable or cationic lipid; 5 to 30 mol % of a non-cationic lipid; 20 to 50 mol % of cholesterol or a derivative thereof; and 1 to 10 mol % of a polymer-lipid. In certain embodiments, the LNP comprises 40 to 60 mol % of an ionizable or cationic lipid; 5 to 15 mol % of a neutral or Zwitterionic lipid; 30 to 50 mol % of cholesterol or a derivative thereof; and 2 to 4 mol % of a polymer-lipid. The mol % of each of the ionizable or cationic lipid, a non-cationic lipid, a polymer-lipid, and sterol lipid are determined together ("total lipid") irrespective of any cargo or excipients.

In some embodiments, the LNP comprises 20-60 mol% ionizable lipid, 5-25 mol% non-cationic lipid, 25-55 mol% sterol, and 0.5-15 mol% PEG-lipid. In some embodiments, the LNP comprises 40-50 mol% ionizable amino lipid, 5-15 mol% non-cationic lipid, 20-40 mol% cholesterol, and 0.5-3 mol% PEG-lipid. In some embodiments, the LNP comprises 45-50 mol% ionizable amino lipid, 9-13 mol% non-cationic lipid, 35-45 mol% cholesterol, and 2-3 mol% PEG-lipid. In some embodiments, the LNP comprises 48 mol% ionizable amino lipid, 11 mol% neutral lipid, 68.5 mol% cholesterol, and 2.5 mol% PEG-lipid.

The LNP may be characterized by a molar ratio or mass ratio of a cargo (such as an exogenous mRNA or oligomeric agent) to total lipid. In certain embodiments, the lipid to cargo ratio is 20:1 to 1:1, 10:1 to 1:1, or 5:1 to 1:1 (mass:mass).

In certain embodiments, the LNP has a mean diameter in a range from X nm to Y nm, wherein X represents the lowest number in the range and Y represents the highest number in the range, wherein X and Y are each independently selected from 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46,47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 518, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200; provided that $X \leq Y$.

In certain embodiments, the LNP has a mean diameter of from 20 nm to 150 nm, from 30 nm to 100 nm, from 40 nm to 150 run, from 50 nm to 150 nm, from 60 nm to 130 nm, from 70 nm to 110 nm, from 70 to 90 nm, from 30 nm to 60 nm, or from 50 nm to 60 nm. The size of LNPs can be measured using a variety

of techniques known in the art such as dynamic light scattering, size exclusion chromatography, nuclear magnetic resonance spectroscopy, and/or microscopy.

A polydispersity index may be used to indicate the particle size distribution of a population of LNPs. A small (e.g., less than 0.3) polydispersity index generally indicates a narrow particle size distribution. In certain embodiments, a polydispersity index (PDI) of diameter of a population of LNPs is less than 2 or less than 1. In certain embodiments, a polydispersity index of diameter of a population of LNPs is 0 to 0.25.

The surface charges of LNPs are related to lipid head groups. The surface potential, which depends on the surface charge density, affects the interactions between particles and the adsorption of counterions. Surface-charged particles repel each other, which is believed to slow or prevent aggregation, and is believed to affect stability of the LNP. Surface charge of LNPs is most often expressed by a zeta potentials, which is the electrical potential of a particle measured from a plane just outside the layer of fluid bound to the particle. Zeta potential is commonly calculated from electrophoretic mobility. LNP compositions with relatively low charges at physiological pH, positive or negative, are generally believed to be desirable. In some embodiments, the zeta potential of the LNP may be from about -10 mV to about +20 mV, from about -5 mV to about +5 mV.

In general, an LNP is a metastable system comprising a distal surface that contacts a surrounding medium and an internal cavity which carries the cargo. In certain embodiments, the LNP comprises a lipid monolayer or bilayer which may encompass a portion of the LNP, e.g., some or all of the distal surface of the LNP. The LNP may comprise a lipid monolayer or bilayer in an internal cavity, which optionally forms multiple vesicles in a single particle. The LNP may comprise a mixture of tightly associated lipids and cargo. The half life of an LNP *in vitro* in aqueous suspension may be on the order of, e.g., one day, ten days, 100 days, 1000 days, or 10,000 days. The LNP may comprise solid phase lipids, liquid phase lipids, or liquid-crystalline lipids, or mixtures thereof. In certain embodiments, the LNP may be formulated as a nanoparticle such as a nucleic acid-lipid nanoparticle described in WO 2009/127060.

In certain embodiments, the target organs to which a cargo is delivered by the LNP include, but are not limited to the liver, lung, heart, spleen, as well as to tumors. In certain embodiments, the target organ is liver. In certain embodiments, the target cell to which a cargo is delivered by the LNP is a hepatocyte and/or a liver sinusoidal endothelial cell. In certain embodiments, the target cell may be *in vivo*, *ex vivo* or *in vitro*.

a. Ionizable and Cationic Lipids

In certain embodiments, provided is an LNP comprising an ionizable or cationic lipid. A lipid is generally a hydrophobic moiety, which may comprise a hydrocarbon chain.

It is believed that design of an ionizable lipid may enhance endosomal drug escape. Thus, it is believed that, *in vivo*, delivery capacity of an LNP may benefit from electrostatic charge interactions with the

endosome and subsequent LNP destabilization to release the cargo. It is believed that an ionizable lipid can aid delivery by either (i) protonating in mildly acidic environments and acquire positive charges and/or (ii) incorporating pH-labile groups that cleave or change conformation in a low pH environment. Such pH-sensitive lipids therefore help enhance the charge-based uptake of liposomes in target cells or trigger cargo release by destabilizing liposome membranes.

In certain embodiments, the ionizable lipid comprises an ionizable head group. In some embodiments, the ionizable head group is a tertiary amine head group.

In certain embodiments, the ionizable lipid comprises a hydrophobic tail. In some embodiments, the hydrophobic tail has two fatty acid chains.

In some embodiments, the head group and the tail are joined by a linker group.

Fusogenicity is a property which describes the ability of a lipid or multi-lipid construct to join with a lipid layer. See, e.g., Hashiba *et al.*, *Small Science*. 2023, *3*(1), 2200071; V. Gyanani, and R. Goswami, *Pharmaceutics*. 2023, *15*(4): 1184. Lipid designs that are believed to impart fusogenic character to a lipid are (i) unsaturation in the lipid tails, (ii) branching in the lipid carbon chains, (iii) multi-tail lipids, e.g., 7C1 and G0-C14 lipids, see, e.g., Dahlman *et al.*, *Nat Nanotechnol.* 2014, *9*(8): 648–655, and (iv) incorporation of polymer-lipids.

It is believed that ionizable or cationic lipids comprising hydrocarbon chains including unsaturation may provide LNPs having higher fluidity. Example publications describing ionizable and/or cationic lipids include U.S, Patent Publication Nos. 2006/0083780 and 2006/0240554; U.S. Pat. Nos. 5,208,036; 5,264,618; 5,279,833; 5,283,185; 5,753,613; and 5,785,992; and PCT Publication No. WO 96/10390, the disclosures of each of which are herein incorporated by reference in their entirety.

Exemplary ionizable and/or cationic lipids may include an ionizable or cationic head group, optionally including one or more linear hydrocarbon chains of 10-20 carbon atoms, and optionally one or more branched, optionally unsaturated, hydrocarbon chains of 12-30 carbon atoms, each of which is optionally interrupted by one or more functionalities selected from ester, ether, amine (e.g., tertiary amine), amide, carbonate, carbamate, urea, and disulfide. The ionizable or cationic lipid may comprise a branching moiety. In certain embodiments the branching moiety comprises a tertiary carbon atom, a quaternary carbon atom, a tertiary amine, a vicinal diol, an amide, a carbamate, or an acetal.

In certain embodiments, the ionizable or cationic lipid is an ionizable lipid containing an amine ("ionizable amino lipid"), for example, a tertiary amine. It is believed that proton cycling of an amino group at disparate pH may provide additional stabilization of a negatively charged cargo (e.g., an exogenous mRNA), while promoting cargo release inside a low pH compartment in vivo (e.g., an endosome).

In certain embodiments, the ionizable lipid has a pKa of the ionizable (e.g., amino) group in the range of about 4 to about 7. Such ionizable lipids have a positive charge in an acidic buffer, thereby allowing for

efficient encapsulation of a negatively charged nucleic acid during formulation, and wherein upon *in vivo* administration at physiological pH, the ionizable lipid is neutral. In certain embodiments, the pKa is in a range of X to Y, wherein X represents the lowest number in the range and Y represents the highest number in the range, wherein X and Y are each independently selected from 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9 and 7.0; provided that X≤Y. In certain embodiments, the pKa is in a range of 6.2-6.6, or 6.6-6.9 or 6.2-6.9. Because such lipids will be largely surface neutralized at physiological pH, it is believed that this will reduce susceptibility to clearance. pKa measurements of lipids within lipid particles can be performed, for example, by using the fluorescent probe 2-(p-toluidino)-6-napthalene sulfonic acid (TNS), using methods described in Cullis *et al.*,*Chem Phys Lipids.* 1986, *40*, 127-144. See also Hassett et al. (2019) *Mol Ther Nucleic Acids* 2019 Apr 15:15:1-11.

In certain embodiments, the ionizable or cationic lipid may comprise a single amine group. Such lipids are described herein and known in the art, e.g., as described in WO 2013/126803, WO 2021/163339, WO 2021/188389, and WO 2022/060871, the disclosure of each of which is herein incorporated by reference in its entirety.

In certain embodiments, the ionizable or cationic lipid may comprise a plurality of amine groups. Such lipids are described herein and known in the art, e.g., as described in WO 2010/129709, WO 2015/199952, WO 2018/191657, WO 2019/036008, WO 2019/036030, WO 2019/036030, WO 2019/036000, WO 2020/081938, and WO 2020/146805, the disclosure of each of which is herein incorporated by reference in its entirety.

In certain embodiments, the ionizable or cationic lipid is l,2-dilinoleyloxy-N,N-dimethylaminopropane (DLinDMA), 1,2- dilinolenyloxy-N,N-dimethylaminopropane (DLenDMA), 2,2-dilinoleyl-4-(2-dimethylaminoethyl)-[l,3]-dioxolane (DLin-K-C2-DMA;"XTC2"), 2,2-dilinoleyl-4-(3-dimethylaminopropyl)-[1,3]-dioxolane (DLin-K-C3-DMA), 2,2-dilinoleyl-4-(4-dimethylaminobutyl)-[l,3]-dioxolane (DLin-K-C4-DMA), 2,2-dilinoleyl-5-dimethylaminomethyl-[1,3]-dioxane (DLin-K6-DMA), 2,2-dilinoleyl-4-N-methylpepiazino-[l,3]-dioxolane (DLin-K-MPZ), 2,2-dilinoleyl-4-dimethylaminomethyl-[l,3]-dioxolane (DLin-K-DMA), l,2-dilinoleylcarbamoyloxy-3-dimethylaminopropane (DLin-C-DAP), 1,2-dilinoleyoxy-3-(dimethylamino)acetoxypropane (DLin-DAC), 1,2-dilinoleyoxy-3-morpholinopropane (DLin-MA), l,2-dilinoleoyl-3-dimethylaminopropane (DLinDAP), 1,2-dilinoleylthio-3-dimethylaminopropane (DLin-S-DMA), 1-linoleoyl-2-linoleyloxy-3-dimethylaminopropane (DLin-2-DMAP), l,2-dilinoleyloxy-3-trimethylaminopropane chloride salt (DLin-TMA.Cl), l,2-dilinoleoyl-3-trimethylaminopropane chloride salt (DLin-TAP.Cl), 1,2-dilinoleyloxy-3-(N-methylpiperazino)propane (DLin-MPZ), 3-(N,N-dilinoleylamino)-1,2-propanediol (DLinAP), 3-(N,N-dilinoleylamino)-l,2-propanediol (DOAP), 1,2-dilinoleyloxo-3-(2-N,N-dimethylamino)ethoxypropane (DLin-EG-DMA), N,N-dioleyl-N,N-dimethylammonium chloride (DODAC), l,2-dioleyloxy-N,N-dimethylaminopropane (DODMA), 1,2-distearyloxy-N,N-dimethylaminopropane

(DSDMA), N-(1-(2,3-dioleyloxy)propyl)-N,N,N-trimethylammonium chloride (DOTMA), N,N-distearyl-N,N-dimethylammonium bromide (DDAB), N-(l-(2,3-dioleoyloxy)propyl)-N,N,N-trimethylammonium chloride (DOTAP), 3-(N-(N',N'-dimethylaminoethane)-carbamoyl)cholesterol (DC-Chol), N-(1,2-dimyristyloxyprop-3-yl)-N,N-dimethyl-N-hydroxyethyl ammonium bromide (DMRIE), 2,3-dioleyloxy-N-[2(spermine-carboxamido)ethyl]-N,N-dimethyl-l-propanaminiumtrifluoroacetate (DOSPA), dioctadecylamidoglycyl spermine (DOGS), 3-dimethylamino-2-(cholest-5-en-3-beta-oxybutan-4-oxy)-1-(cis,cis-9,12-octadecadienoxy)propane (CLinDMA), 2-[5'-(cholest-5-en-3-beta-oxy)-3'-oxapentoxy)-3-dimethyl-l-(cis,cis-9',l-2'-octadecadienoxy)propane (CpLinDMA), N,N-dimethyl-3,4-dioleyloxybenzylamine (DMOBA), 1,2-N,N'-dioleylcarbamyl-3-dimethylaminopropane (DOcarbDAP), or 1,2-N,N'-dilinoleylcarbamyl-3-dimethylaminopropane (DLincarbDAP), or mixtures thereof. In certain embodiments, the ionizable lipid is DLinDMA, DLin-K-C2-DMA ("XTC2"), or mixtures thereof.

In certain embodiments, the ionizable or cationic lipid is N,N -dioleyl-N,N -dimethylammonium chloride (DODAC), 1 ,2-dioleyloxy-N,N- dimethylaminopropane (DODMA), 1,2-distearyloxy-N,N-dimethylaminopropane (DSDMA), N- (l-(2,3-dioleyloxy)propyl)-N,N,N-trimethylammonium chloride (DOTMA), N,N-distearyl-N,N- dimethylammonium bromide (DDAB), N-(l-(2,3-dioleoyloxy)propyl)-N,N,N- trimethylammonium chloride (DOTAP), 3-(N— (N',N'-dimethylaminoethane)-carbamoyl)cholesterol (DC -Choi), N-(1 ,2-dimyristyloxyprop-3-yl)-N,N-dimethyl-N- hydroxyethyl ammonium bromide (DMRIE), 2,3-dioleyloxy-N-[2(spermine- carboxamido)ethyl]-N,N-dimethyl-l-propanaminiumtrifluoroacetate (DOSPA), dioctadecylamidoglycyl spermine (DOGS), 3-dimethylamino-2-(cholest-5-en-3-beta-oxybutan- 4-oxy)-l-(cis,cis-9,12-octadecadienoxy)propane (CLinDMA), 2-[5'-(cholest-5-en-3.beta.-oxy)- 3 '-oxapentoxy)-3 -dimethyl- 1 -(cis,cis-9', 1 -2'-octadecadienoxy)propane (CpLinDMA), N,N- dimethyl-3,4-dioleyloxyben2ylamine (DMOBA), 1 ,2-N,N'-dioleylcarbamyl-3- dimethylaminopropane (DOcarbDAP), 1 ,2-N,N'-Dilinoleylcarbamyl-3-dimethylaminopropane (DLincarbDAP), 1,2-Dilinoleoylcarbamyl-3-dimethylaminopropane (DLinCDAP), and mixtures thereof. A number of these lipids and related analogs have been described in U.S. Patent Publication Nos. 20060083780 and 20060240554; U.S. Pat. Nos. 5,208,036; 5,264,618; 5,279,833; 5,283,185; 5,753,613; and 5,785,992; and PCT Publication No. WO 96/10390, the disclosures of which are each herein incorporated by reference in their entirety.

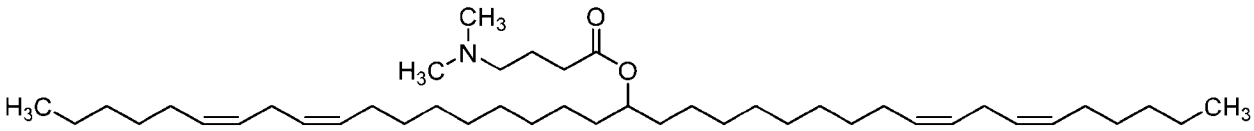
In certain embodiments, the ionizable or cationic lipid comprises multiple sites of protonation. See, e.g., Qin et al., Signal Transduction and Targeted Therapy (2022) 7:166, which is incorporated by reference herein in its entirety. In certain embodiments, the ionizable or cationic lipid comprises 2, 3, 4, 5 or 6 sites of protonation (e.g., 2, 3, 4, 5, or 6 amino groups). See, e.g., WO 2010/053572.

Further ionizable or cationic lipids include those disclosed in, e.g., WO 2011/068810, WO 2012/000104, WO 2012/170930, WO 2013/086354, WO 2018/006052, WO 2020/097520, WO 2021/000041,

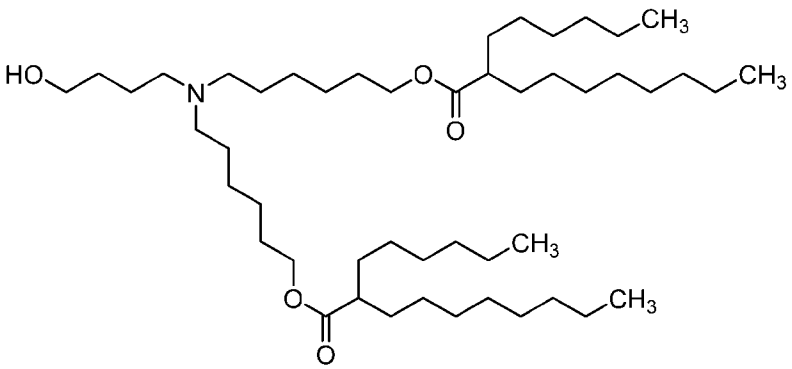
WO 2022/266032, WO 2021/026647, WO 2022/173531, the disclosure of each of which is herein incorporated by reference in its entirety.

An ionizable or cationic lipid can be synthesized according to methods known in the art. The synthesis of lipids such as DLin-K-C2-DMA ("XTC2"), DLin-K-C3-DMA, DLin-K-C4-DMA, DLin-K6-DMA, and DLin-K-MPZ, as well as other ionizable or cationic lipids, is described in WO 2010/042877, the disclosure of which is herein incorporated by reference in its entirety. The synthesis of lipids such as DLin-K-DMA, DLin-C-DAP, DLin-DAC, DLin-MA, DLinDAP, DLin- S-DMA, DLin-2-DMAP, DLin-TMA.Cl, DLin-TAP.Cl, DLin-MPZ, DLinAP, DOAP, and DLin-EG-DMA, as well as other lipids, is described in WO 2009/086558, the disclosure of which is herein incorporated by reference in its entirety. The synthesis of lipids such as CLinDMA, as well as other lipids, is described in U.S. Patent Publication No. 2006/0240554, the disclosure of which is herein incorporated by reference in its entirety.

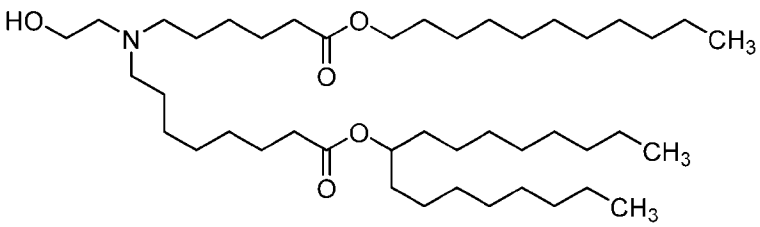
In certain embodiments, the ionizable or cationic lipid is DLin-MC3-DMA, ALC-0315, SM-102, or LP-01:



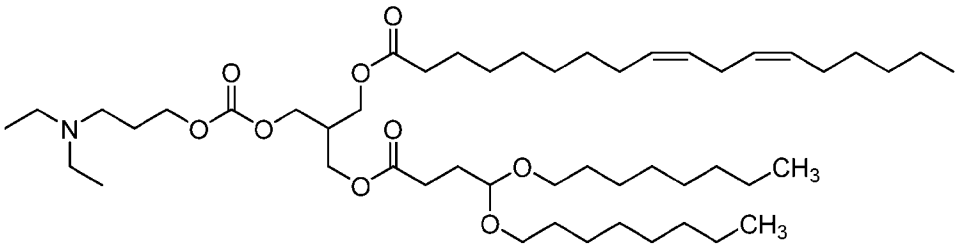
DLin-MC3-DMA



ALC-0315



SM-102



LP-01.

b. Polymer-Lipids

Polymer lipids such as PEG lipids may be used as a component in an LNP to modify the surface of the LNP. Such modification can stabilize an LNP in the presence of serum and therefore assist in extending circulation *in vivo* due to reduced protein absorption. In addition, the amount of polymer lipid in the LNP influences overall LNP size, which can affect the rate of cellular uptake.

In certain embodiments provided is an LNP comprising a polymer-lipid. In certain embodiments, the polymer lipid is a pegylated or PEG lipid. When formulated as part of an LNP, PEG lipids can mask or cloak the cargo *in vivo*, thereby reducing immunogenicity and antigenicity of the cargo.

The polymer-lipid may be a polymer-functionalized lipid, in which the polymer and lipid (e.g., a hydrocarbon chain optionally interrupted by one or more intervening functionalities) are joined by covalent bonds with optional intervening atoms. The polymer-lipid optionally includes a branching moiety. The lipid may be, e.g., a hydrocarbon chain optionally interrupted by one or more intervening functionalities. The polymer-lipid generally includes an uncharged, hydrophilic moiety which is believed to limit aggregation, such as PEG, GMI, or ATTA, during formulation of an LNP. Thus, it is believed that a polymer-lipid may also reduce aggregation when included in an LNP. In certain embodiments, the content of the polymer-lipid in the LNP is selected to reduce particle aggregation.

Examples of polymer-lipids include polyethylene glycol (PEG)-modified lipids, monosialoganglioside GMI, and polyamide oligomers ("PAO") such as described in U.S. Pat. No. 6,320,017. ATTA-lipids are described, e.g., in U.S. Patent No. 6,320,017, and PEG-functionalized lipid are described, e.g., in U.S. Patent Nos. 5,820,873, 5,534,499 and 5,885,613. In certain embodiment, the polymer-lipid comprises one or more hydrocarbon chains that are interrupted by a biodegradable functional group, e.g., an ester. The polymer-lipid may comprise a branching moiety. In certain embodiments the branching moiety comprises a tertiary carbon atom, a quaternary carbon atom, a tertiary amine, a vicinal diol, an amide, a carbamate, or an acetal.

The lipid and optional branching moiety are believed to influence associative strength of a polymer-lipid with the LNP. It is believed that at least three characteristics influence the rate of exchange: length of lipid chain, rigidity (e.g., as determined by saturation) of lipid chain, and size of the steric-barrier head group (see, e.g., U.S. Patent No. 5,820,873). For some therapeutic applications it may be preferable for the PEG-

functionalized lipid to be rapidly lost from the LNP in vivo, and thus the PEG-lipid would have relatively short lipid anchors. In other therapeutic applications it may be preferable for a nucleic acid-lipid particle to exhibit a longer plasma circulation lifetime and hence the PEG-lipid will possess relatively longer lipid anchors. It is believed that mPEG (mw2000)-diastearoylphosphatidylethanolamine (PEG-DSPE) will remain associated with an LNP for days in vivo. Other conjugates, such as PEG-CerC20 have similar staying capacity. PEG-CerC14, however, is believed to more rapidly exchange out of the formulation upon exposure to serum.

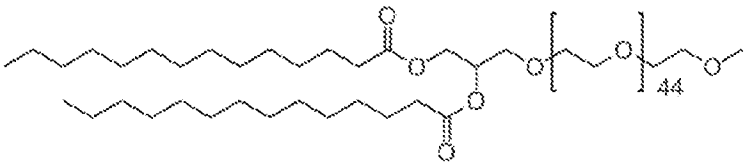
In certain embodiments, the polymer-lipid is non-cationic. In certain embodiments, the non-cationic lipid is charge neutral. In certain embodiments, the polymer-lipid is Zwitterionic.

The polymer may be a hydrophilic polymer, for example, a poly(ethylene glycol), also referred to as a poly(ethylene oxide) or PEG ("PEG-lipid"). In certain embodiments, PEG is an optionally substituted linear polymer of ethylene glycol or ethylene oxide. In certain embodiments, the PEG moiety is substituted, e.g., by one or more alkyl, alkoxy, acyl, hydroxy, or aryl groups. In certain embodiments, the PEG moiety includes PEG copolymer such as PEG-polyurethane or PEG-polypropylene (see, e.g., J. Milton Harris, Poly(ethylene glycol) chemistry: biotechnical and biomedical applications (1992)). In certain embodiments, the PEG moiety is a PEG homopolymer.

Examples of polymer lipids include PEG-functionalized phosphatidylethanolamine and phosphatidic acid, PEG-ceramide conjugates (e.g., PEG-CerC14 or PEG-CerC20 which are described in U.S. Patent No. 5,820,873), PEG-modified dialkylamines, PEG-modified 1,2-diacyloxypropan-3-amines, PEG-modified diacylglycerols, and PEG-modified dialkylglycerols. The lipid chain may vary according to known determinants in the art and may be, for example, a hydrocarbon of 10 to 30 carbon atoms in length.

For example, certain embodiments provide a pegylated diacylglycerol (PEG-DAG) such as 1-(monomethoxy-polyethyleneglycol)-2,3-dimyristoylglycerol (PEG-DMG), PEG-dipalmitoylglycerol, PEG-dilauroylglycerol, PEG-distearoylglycerol (PEG-DSPE), or 4-O-(2',3'-di(tetradecanoyloxy)propyl-1-O-(ω-methoxy(polyethoxy)ethyl)butanedioate (PEG-S-DMG); a pegylated phosphatidylethanolamine (PEG-PE); a PEG-glycamide ((N-acyl-N-alkyl-glycamine) such as PEG- dimyristylglycamide, PEG-dipalmitoylglycamide, or PEG-disterylglycamide; a PEG dialkoxypropylcarbamate such as ω-methoxy(polyethoxy)ethyl-N-(2,3-di(tetradecanoxy)propyl)carbamate or 2,3-di(tetradecanoxy)propyl-N-(ω-methoxy(polyethoxy)-ethyl)carbamate; PEG-cholesterol; PEG-dialkyloxypropyls (e.g., PEG-DAA), PEG-phosphatidylethanolamine, and PEG-ceramide (see, e.g., U.S. Pat. No. 5,885,613). For example, a polymer lipid may be PEG-DMG, PEG-c-DOMG, PEG-DLPE, PEG-DMPE, PEG-DPPC, or a PEG-DSPE, or a combination thereof.

In general, the length of the PEG chain may be represented by either a number indicating the number of repeating ethylene oxide units, or by a molecular weight in Daltons of the PEG portion of the polymer-lipid. In a particular example, PEG-DMG 2000 has the following structure:



where 2000 is the approximate molecular weight in Daltons of the PEG portion of the molecule. The length of PEG chain may vary between individual lipids in a sample and thus may be expressed as an average number of repeating units or average molecular weight.

A polymer lipid such as a PEG-lipid described herein may comprise a methyl (methoxy) or a hydroxyl group at the terminus of the PEG chain. In certain embodiments, the polymer lipid comprises an methoxy group at the terminus of a PEG chain. In certain embodiments, the polymer lipid is a PEG-OH lipid and comprises an -OH group at the terminus of a PEG chain. A polymer lipid can comprise a defined number of ethylene glycol units, for example, at least 1, at least 2, at least 5, at least 10, 10-150, or 40-60 ethylene glycol units. In certain embodiments, a number average molecular weight of the polymer portion of a polymer lipid is from about 200 Da to about 5000 Da.

In certain embodiments, the polymer lipid has a molecular weight from about 1500 Da to about 3500 Da.

In certain embodiments, the LNP comprises two or more polymer lipids having distinct structures.

In general, the length of the PEG chain may be represented by either a number indicating the number of repeating ethyl oxide units, or by a molecular weight in Daltons of the PEG portion of the polymer-lipid.

c. Non-Cationic Lipids

Non-cationic lipids can help stabilize an LNP and form the basic outer layer structure of an LNP. In certain embodiments, the LNP comprises a non-cationic lipid. In certain embodiments, the non-cationic lipid is a (charge) neutral or Zwitterionic lipid. The neutral or Zwitterionic lipid can generally be any lipid species which is uncharged or neutral zwitterionic form at physiological pH. Generally, the non-cationic lipid will include a polar head group and one or more (e.g., two) hydrophobic tail groups. Exemplary head groups include phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and inositol.

In certain embodiments, the neutral lipid comprises two hydrocarbon groups which are each optionally interrupted with a biodegradable moiety. Non-cationic lipids having a variety of acyl chain groups of varying chain length and degree of saturation are available or may be isolated or synthesized by well-known techniques. In certain embodiments, the non-cationic lipid comprises saturated fatty acids, or mono- or di-unsaturated fatty acids. Additionally, lipids having mixtures of saturated and unsaturated fatty acid chains can be used. In certain embodiments, a fatty acid is interrupted by a biodegradable moiety such as an ester.

In certain embodiments, the non-cationic lipid comprises a phosphatidylcholine (PC), phosphatidylethanolamine (PE), glycerophospholipid, sphingophospholipid, sphingolipid, phosphono lipids, natural lecithins, or hydrogenated phospholipid. Such lipids include, for example diacylphosphatidylcholine,

diacylphosphatidylethanolamine, ceramides, sphingomyelin, dihydrosphingomyelin, cephalin, and cerebrosides.

The selection of non-cationic lipid is generally guided by consideration of, e.g., LNP particle size and stability in circulation.

In certain embodiments, the non-cationic lipid is a phospholipid. As used herein, a "phospholipid" refers to a lipid that includes a hydrophilic phosphate head group and one or more hydrophobic tail groups. In some embodiments, a phospholipid may facilitate fusion to a membrane. For example, a cationic phospholipid may interact with one or more negatively charged phospholipids of a membrane (e.g., a cellular or intracellular membrane). Fusion of a phospholipid to a membrane may allow delivery of the one or more components of the LNP, e.g., the cargo, through the membrane, e.g., into a cell.

In certain embodiments, the phospholipid is a phosphatidylcholine. Exemplary phosphatidylcholines include, but are not limited to, 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), dipalmitoyl phosphatidylcholine, dipalmitoyl -sn-glycero-3-phosphocholine (DPPC), 2-oleoyl-1-palmitoyl-sn-glycero-3-phosphocholine (POPC), dimyristoyl phosphatidylcholine (DMPC), and dioleoyl phosphatidylcholine (DOPC).

In certain embodiments, the phospholipid is a phosphatidylethanolamine. In certain embodiments, the phosphatidylethanolamine is distearoyl phosphatidylethanolamine (DSPE), dipalmitoyl phosphatidyl ethanolamine (DPPE), 1,2- dioleoyl-sn-glycero-3- phosphoethanolamine (DOPE), dimyristoyl phosphoethanolamine (DMPE), 16:0-monomethyl phosphatidylethanolamine, 16:0-dimethyl phosphatidylethanolamine, 18:1-trans phosphatidylethanolamine, palmitoyl oleoyl-phosphatidylethanolamine (POPE), or 1-stearoyl-2-oleoyl-phosphatidyl ethanolamine (SOPE).

In certain embodiments, the phospholipid comprises a glycerophospholipid. In certain embodiments, the glycerophospholipid is plasmalogen, phosphatidate, or phosphatidylcholine. In certain embodiments, the glycerophospholipid is phosphatidylserine, phosphatidic acid, phosphatidylglycerol, phosphatidylinositol, palmitoyl oleoyl phosphatidylglycerol (POPG), or lysophosphatidylcholine. In some embodiments, the phospholipid comprises a sphingophospholipid. In some embodiments, the sphingophospholipid is sphingomyelin, ceramide phosphoethanolamine, ceramide phosphoglycerol, or ceramide phosphoglycerophosphoric acid.

In certain embodiments, the phospholipid comprises a natural membrane lipid, e.g. a lecithin. In some embodiments, the natural lecithin is egg yolk lecithin or soybean lecithin. In some embodiments, the phospholipid comprises a hydrogenated phospholipid. In some embodiments, the hydrogenated phospholipid is hydrogenated soybean phosphatidylcholine.

In certain embodiments, the non-cationic lipid is selected from 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE); 1,2-dilinoleoyl- sn-glycero-3-phosphocholine (DLPC); 1,2-dimyristoyl-sn-

glycero-phosphocholine (DMPC); 1,2 dioleoyl-sn-glycero-3-phosphocholine (DOPC); 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC); 1,2-diundecanoyl-sn-glycero-phosphocholine (DUPC), distearoylphosphatidylcholine (DSPC), dioleoylphosphatidylcholine (DOPC), dipalmitoylphosphatidylcholine (DPPC), 1,2-dierucoyl-sn-glycero-3-phosphocholine (DEPC), 2,3-dipalmitoyl-sn-glycero-1-phosphocholine, dioleoylphosphatidylglycerol (DOPG), dipalmitoylphosphatidylglycerol (DPPG), dioleoylphosphatidylethanolamine (DOPE), palmitoyloleoylphosphatidylcholine (POPC), palmitoyloleoylphosphatidylethanolamine (POPE) and diolcoylphosphatidylcthanolaminc 4-(N-malcimidomcthyl)-cyclohcxanc-1-carboxylatc (DOPE-mal), dipalmitoyl phosphatidyl ethanolamine (DPPE), dimyristoylphosphoethanolamine (DMPE), distearoyl-phosphatidylethanolamine (DSPE), 16:0-monomethyl phosphatidylethanolamine, 16:0-dimethyl phosphatidylethanolamine, 18:1-trans phosphatidylethanolamine, 1-stearioyl-2-oleoylphosphatidyethanol amine (SOPE), 1,2-dielaidoyl-sn-glycero-3- phophoethanolamine (transDOPE), 1,2-dioctadecenyl-sn-glycero-3-phosphocholine (18:0 diether PC); l-oleoyl-2-cholesterylhemisuccinoyl-sn-glycero-3-phosphocholinc (OChcmsPC); 1-hcxadccyl-sn- glyccro-3-phosphocholinc (CI 6 Lyso PC); 1,2-dilinolcnoyl-sn-glycero-3- phosphocholine; 1,2-diarachidonoyl-sn-glycero-3-phosphocholine; 1,2-didocosahexaenoyl-sn-glycero-3-phosphocholine; 1,2-diphytanoyl-sn-glycero-3- phosphoethanolamine (ME 16.0 PE); 1,2-distearoyl-sn-glycero-3-phosphoethanolamine; 1,2-dilinoleoyl-sn-glycero-3-phosphoethanolamine; 1,2-diarachidonoyl-sn-glycero-3-phosphoethanolamine; 1,2-didocosahexaenoyl-sn-glycero-3-phosphoethanolamine; 1,2-dioleoyl-sn-glycero-3-phospho-rac-(1-glycerol) sodium salt (DOPG); sphingomyelins (SM); and ceramides. Such lipids may be synthetic or naturally derived.

In certain embodiments, the non-cationic lipid is selected from DSPC, DPPC, DMPC, DOPC, POPC, DOPE and SM.

In certain embodiments, the non-cationic lipid is 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC).

In certain embodiments, the LNP comprises a plurality of non-cationic lipids, for example, 2, 3, or 4 distinct phospholipids selected from those described herein or known in the art.

d. Sterols and derivatives thereof

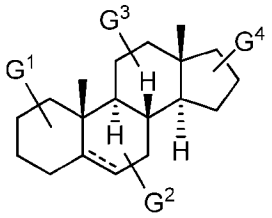
Sterols can also help stabilize an LNP. In certain embodiments the sterol is cholesterol.

Cholesterol is a ubiquitous structural membrane lipid and its role in a lipid composition depends on context. When combined with phospholipids with low gel-liquid crystalline phase transitions (Tm), cholesterol is believed to aid formation of a liquid-ordered phase which is characterized by increased bilayer thickness and membrane rigidity. It is believed that cholesterol and other linear and/or branched hydrocarbon-containing (e.g., low Tm) lipids combine such that the cross-sectional area of the lipid and cholesterol is

lower than the sum of the individual cross-sectional areas. However, when combined with high Tm lipids, cholesterol is believed to increase membrane fluidity and provide narrower bilayers. In any case, cholesterol is believed to encourage a liquid-ordered phase. Additionally, cholesterol is believed to reduce the amount of surface-bound protein and improve circulation half-life. The amount of cholesterol in an LNP may be selected to match the in vivo membrane cholesterol content.

In some embodiments, provided is an LNP comprising cholesterol.

Many isomers and derivatives of cholesterol are present in organisms throughout nature. In certain embodiments, the LNP comprises a cholesterol derivative. In certain embodiments, the cholesterol derivative has the formula:



wherein G^1, G^2, G^3, and G^4 are each independently 1-4 substituent(s) selected from halo, cyano, hydroxy, C_{1-6}alkyl optionally substituted with R^{a1}, C_{1-6}haloalkyl, C_{1-6}hydroxyalkyl, C_{1-6}heteroalkyl, C_{3-10}cycloalkyl optionally substituted with R^{a1}, C_{3-10}cycloalkyl-C_{1-6}alkyl optionally substituted with R^{a1}, C_{6-10}aryl optionally substituted with R^{a1}, C_{6-10}aryl-C_{1-6}alkyl optionally substituted with R^{a1}, heteroaryl optionally substituted with R^{a1}, heteroaryl-C_{1-6}alkyl optionally substituted with R^{a1}, heterocyclyl optionally substituted with R^{a1}, heterocyclyl-C_{1-6}alkyl optionally substituted with R^{a1}, OR^{a2}, $-NH_2$, $-NHR^{a2}$, $-N(R^{a2})_2$, $-C_{1-6}$alkylene-NH_2, $-C_{1-6}$alkylene-NHR^{a2}, $-C_{1-6}$alkylene-$N(R^{a2})_2$, $-C(O)R^{a3}$, $-C(O)OR^{a3}$, $-C(O)NHR^{a3}$, $-C(O)N(C_{1-4}alkyl)R^{a3}$, $-S(O)_2R^{a3}$, $-S(O)R^{a3}$, $-NHC(O)R^{a3}$, $-N(C_{1-4}alkyl)C(O)R^{a3}$, $-NHS(O)R^{a3}$, $-N(C_{1-4}alkyl)S(O)R^{a3}$, $-NHS(O)_2R^{a3}$, and $-N(C_{1-4}alkyl)S(O)_2R^{a3}$; each R^{a2} is independently selected from C_{1-6}alkyl, C_{3-10}cycloalkyl, C_{6-10}aryl, heteroaryl, and heterocyclyl; each R^{a3} is independently hydrogen, -OH, C_{1-6}alkyl, C_{1-6}haloalkyl, C_{3-10}cycloalkyl, C_{6-10}aryl, heteroaryl, or heterocyclyl; each R^{a1} is independently halo, cyano, hydroxy, $-NH_2$, $-NHR^{a4}$, $-N(R^{a4})_2$, C_{1-6}alkyl, C_{1-6}haloalkyl, OR^{a4}, or C_{3-10}cycloalkyl; each R^{a4} is independently selected from C_{1-6}alkyl, C_{3-10}cycloalkyl, C_{6-10}aryl, heteroaryl, and heterocyclyl, and each R^{a4} is optionally substituted with hydroxy, one to six halo, or C_{1-3}alkoxy. In certain embodiments, the cholesterol derivative is selected from β-sitosterol, β-sitosterol acetate, a β-sitosterol amino acid conjugate, fecosterol, ergosterol, 9,11-dehydroergosterol, campesterol, stigmasterol, brassicasterol, fucosterol, tomatidine, ursolic acid, α-tocopherol, daucosterol, cholesterol, 5-heptadecylresorcinol, cholesterol hemisuccinate, 6-keto-5α-hydroxycholesterol, 7α-hydroxycholesterol, 7β-hydroxycholesterol, 7-ketocholesterol, 7β,25-dihydroxycholesterol, 27-hydroxycholesterol, 25-hydroxycholesterol, 20α-hydroxycholesterol, 5α-cholestanol, 5β-coprostanol, cholesteryl-(2'-hydroxy)-ethyl ether, 6-ketocholestanol, cholesteryl-(4'-

hydroxy)-butyl ether, 5α-cholestane, cholestenone, 5β-cholestanone, cholesteryl decanoate, vitamin D3, vitamin D2, calcipotriol, botulin, luperol, ursolic acid, oleanolic acid, DC-cholesterol, BHEM-cholesterol, cholesteryl oleate, or a combination thereof. See, e.g., Paunovska, K. *et al.*, *Adv Mater.* 2019, *31*(14); Patel *et al.*, *Nat. Comm.*, 2020 *11*:983; Ni *et al.*, *Nat. Comm.* 2022, 13, Article number: 4766; Kim *et al.*, *ACS Nano.* 2022, *16*(9), 14792–14806.

In certain embodiments, the cholesterol derivative is a corticosteroid. In certain embodiments, the LNP comprises a corticosteroid selected from cortisone, cortisol, prednisolone, methylprednisolone, 20α-dihydroprednisolone, 20β-dihydroprednisolone, betamethasone, dexamethasone, prednisone, flumethasone, isoflupredone, eclomethasone, clobetasol, triamcinolone acetonide, and hydrocortisone.

2. LNP Preparation

The LNP can be prepared by any method known in the art including, but not limited to, a continuous mixing method or a direct dilution process. LNPs can be generated according to methods known in the art, see for example PCT/US2016/052352; PCT/US2016/068300; PCT/US2017/037551; PCT/US2015/027400; PCT/US2016/047406; PCT/US2016/000129; PCT/US2016/014280; PCT/US2016/014280; PCT/US2017/038426; PCT/US2014/027077; PCT/US2014/055394; PCT/US2016/052117; PCT/US2012/069610; PCT/US2017/027492; PCT/US2016/059575 and PCT/US2016/069491 all of which are incorporated by reference herein in their entirety.

In certain embodiments, provided is a method for preparing an LNP by a continuous mixing method. E.g., a process that includes providing an aqueous solution comprising a nucleic acid in a first reservoir, providing a lipid solution in a second reservoir, and mixing the aqueous solution with the lipid solution such that the organic lipid solution mixes with the aqueous solution so as to rapidly produce a LNP encapsulating the cargo. The lipid solution comprises a lower alcohol such as ethanol. This process and the apparatus for carrying this process are described in detail in U.S. Patent Publication No. 2004/0142025, the disclosure of which is herein incorporated by reference in its entirety. By mixing the aqueous solution comprising a cargo with the organic lipid solution, the organic lipid solution undergoes a continuous stepwise dilution in the presence of the buffer solution (i.e., aqueous solution) to produce an LNP.

In certain embodiments, provided is a method for preparing an LNP by a direct dilution process that includes forming an LNP solution and directly introducing the LNP solution into a collection vessel containing a controlled amount of dilution buffer. The collection vessel may include one or more elements configured to stir the contents of the collection vessel to facilitate dilution. In one aspect, the amount of dilution buffer present in the collection vessel is substantially equal to the volume of liposome solution introduced thereto. As a non-limiting example, a liposome solution in about 45% ethanol when introduced into the collection vessel containing an equal volume of dilution buffer will advantageously yield smaller particles.

In certain embodiments, provided is a method for preparing an LNP by a direct dilution process in which a third reservoir containing dilution buffer is fluidly coupled to a second mixing region. In this embodiment, the LNP solution formed in a first mixing region is immediately and directly mixed with dilution buffer in the second mixing region. In preferred aspects, the second mixing region includes a T-connector arranged so that the LNP solution and the dilution buffer flows meet as opposing 180° flows; however, connectors providing shallower angles can be used. In one aspect, the flow rate of dilution buffer provided to the second mixing region is controlled to be substantially equal to the flow rate of liposome solution introduced thereto from the first mixing region. Such control of the dilution buffer flow rate may advantageously allow for small particle size formation at reduced concentrations. Processes and apparatuses for carrying out direct dilution processes are described in U.S. Patent Publication No. 2007/0042031, the disclosure of which is herein incorporated by reference in its entirety.

The particle size distribution of LNPs can be controlled using manufacturing methods such as extrusion, sonication, homogenization, and microfluidic methods. An LNP provided herein can be sized-adjusted by a method known in the art. One sizing method is described in U.S. Pat. No. 4,737,323, the disclosure of which is herein incorporated by reference in its entirety. Sonicating a particle suspension either by bath or probe sonication may produce a size reduction down to particles of less than about 50 nm in size. Homogenization is another method which relies on shearing energy to fragment larger particles into smaller ones. In a typical homogenization procedure, particles are recirculated through a standard emulsion homogenizer until selected particle sizes are observed.

Particle size distribution can be monitored by conventional methods including laser-beam particle size discrimination, or QELS.

Extrusion of an LNP through a small-pore membrane (e.g., of polycarbonate) or an asymmetric ceramic membrane is also an effective method for reducing particle size or for producing LNPs of low polydispersion. Typically, the suspension is cycled through the membrane one or more times until the desired particle size distribution is achieved. The particles may be extruded through successively smaller-pore membranes, to achieve a gradual reduction in size.

In certain embodiments, the LNP may be formed by a method described in International Publication Nos. WO 2011/127255 or WO 2008/103276, the contents of each of which is herein incorporated by reference in their entirety. In certain embodiments, the LNP may be formulated by the methods described in US Patent Publication No US 2013/0156845 or International Publication No WO 2013/093648 or WO 2012/024526, each of which is herein incorporated by reference in its entirety.

In certain embodiments, an LNP may be formed by a method described in US 9,668,980.

The lipid nanoparticles described herein may be made in a sterile environment, e.g., using the method described in US Patent Publication No. US 2013/0164400, herein incorporated by reference in its entirety. The LNP may be sterilized by sterile filtration.

The efficiency of encapsulation of a cargo describes the amount of cargo that is encapsulated or otherwise associated with a nanoparticle composition after preparation, relative to the initial amount provided. An exemplary method for determining encapsulation efficiency is comparing the amount of cargo in a solution containing the LNP before and after disintegrating the LNP, e.g., using one or more organic solvents or detergents. For example, fluorescence may be used to measure the amount of free cargo (e.g., exogenous mRNA) in a solution. In certain embodiments, the encapsulation efficiency of a cargo may be at least 50%, for example at least 90%.

In certain embodiments, a cargo may be loaded into an LNP following nanoparticle formation. See, e.g., WO 2018/089801.

In certain embodiments, a cargo may be encapsulated in the lipid portion of the LNP or in an aqueous space enveloped by some or all of the lipid portion of the LNP. The encapsulation can be full encapsulation, partial encapsulation, or both. In some embodiments, a cargo is fully encapsulated in the LNP.

In certain embodiments, one or more cargos may be associated with an LNP via a covalent bond or a non-covalent bond. In some embodiments, any of the cargos may be separately or together contained in an LNP.

B. Viral Vectors

In certain embodiments, a nucleic acid encoding an editing system, a nucleic acid binding agent, or a guide, may be introduced into a cell or organism through a viral based or non-viral based delivery system, including adenovirus vectors, adeno-associated virus (AAV) vectors, retrovirus vectors, lentiviral vectors, herpes virus vectors, liposomes, poxviruses, naked DNA administration, plasmids, cosmids, phages, encapsulated cell technology, and the like.

In certain embodiments, a nucleic acid encoding an editing system, a nucleic acid binding agent, or a guide, may be introduced into a cell or organism through a viral vector. In certain embodiments, in some cases, the vector is a plasmid, a minicircle, a CELiD, an adeno-associated virus (AAV) derived virion, or a lentivirus.

Certain Compositions and Methods for Formulating Certain Compositions

In certain embodiments, the guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs, viral vectors and cells disclosed herein are provided as compositions. In certain embodiments, the compositions comprise LNPs. In certain embodiments, the LNPs comprise, encapsulate or are capable of encapsulating a cargo, wherein the cargo is a guided nucleic acid binding agent, a guide, an editing system, a nucleic acid, or an expression system. In certain embodiments, the compositions are pharmaceutical compositions.

Also provided are compositions comprising a plurality of LNPs having distinct compositions and/or cargos. See, e.g., WO 2014/144196, WO 2016/197133. In certain embodiments, provided is a pharmaceutical

composition comprising a first LNP comprising a first cargo and having a composition defined herein, and a second LNP comprising a second cargo and having a composition different from the first LNP, wherein the first cargo and the second cargo are different. In certain embodiments, provided is a pharmaceutical composition comprising a first LNP comprising a first cargo and having a composition defined herein, and a second LNP comprising a second cargo and having a composition substantially identical to the first LNP, wherein the first cargo and the second cargo are different.

In certain embodiments, there is provided a composition comprising a cell, organ or tissue that has been edited or modified by the methods disclosed herein. In certain embodiments, the composition is an autologous or allogeneic cell, organ or tissue that has been edited or modified by the methods disclosed herein. In certain embodiments, the cell, organ or tissue may be *in vitro*. In certain embodiments, the cell may be *ex vivo*.

In certain embodiments, the pharmaceutical composition can have one or more additional reagents, wherein such additional reagents are selected from a buffer, a buffer for introducing a polypeptide or polynucleotide into a cell, a wash buffer, a control reagent, a control vector, a control RNA polynucleotide, a reagent for *in vitro* production of a polypeptide from DNA, adaptors for sequencing and the like. A buffer can be a stabilization buffer, a reconstituting buffer, a diluting buffer, or the like. In some embodiments, the pharmaceutical composition can also include one or more components that can be used to facilitate or enhance on-target binding or cleavage of a target nucleic acid by an endonuclease, or improve specificity of targeting.

In certain embodiments, the pharmaceutical composition may contain one or more excipients to facilitate systemic delivery of an oligonucleotide. In certain embodiments, any components of a pharmaceutical composition are formulated with pharmaceutically acceptable excipients such as carriers, solvents, stabilizers, adjuvants, diluents, etc., depending upon the particular mode of administration and dosage form. Suitable excipients can include, for example, carrier molecules that include large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Other exemplary excipients include antioxidants (for example and without limitation, ascorbic acid), chelating agents (for example and without limitation, EDTA), carbohydrates (for example and without limitation, dextrin, hydroxyalkylcellulose, and hydroxyalkylmethylcellulose), stearic acid, liquids (for example and without limitation, oils, water, saline, glycerol and ethanol), wetting or emulsifying agents, pH buffering substances, and the like.

In certain embodiments, the LNP composition may comprise one or more excipients, for example, a triglyceride, a surfactant, and/or a hydrophobic excipient such as a wax. Triglycerides include trimyristin (Dynasan 114), tripalmitin (Dynasan 116), or tristearin (Dynasan 118), Witeposol bases, glyceryl stearates

(Imwitor 900), glyceryl behenates (Compritol 888 ATO), and glyceryl palmitostearates (Precirol ATO 5). Nonionic surfactants may include moieties such as ethylene glycol esters, propylene glycol esters, glyceryl esters, polyglyceryl esters, sorbitan esters, sucrose esters, and ethoxylated esters, nonionic alkanolamides and ethers such as fatty alcohol ethoxylates, propoxylated alcohols, and ethoxylated/propoxylated block polymers. Anionic surfactants include carboxylates, acyl lactylates, acyl amides of amino acids, esters of sulfuric acid such as alkyl sulfates and ethoxylated alkyl sulfates, sulfonates such as alkyl benzene sulfonates, acyl isethionates, acyl taurates and sulfosuccinates, and phosphates. Specific surfactants include lecithin, Poloxamer 188, Poloxamer 407, Tyloxapol, Polysorbate 20, Polysorbate 60, Polysorbate 80, sodium cholate, sodium glycocholate, taurodeoxycholic acid sodium, butanol, butyric acid, cetylpyridinium chloride, sodium dodecyl sulfate, sodium oleate, polyvinyl alcohol, or Cremophor EL. Waxes include beeswax and cetyl palmitate. Other hydrophobic excipients include stearic acid, palmitic acid, behenic acid, Miglyol 812, and paraffin.

Physiologically tolerable carriers are well known in the art. Exemplary liquid carriers are sterile aqueous solutions that contain no materials in addition to the active ingredients and water, or contain a buffer such as sodium phosphate at physiological pH value, physiological saline or both, such as phosphate-buffered saline. Aqueous carriers can contain more than one buffer salt, as well as salts such as sodium and potassium chlorides, dextrose, polyethylene glycol and other solutes. Liquid compositions can also contain liquid phases in addition to and to the exclusion of water. Exemplary of such additional liquid phases are glycerin, vegetable oils such as cottonseed oil, and water-oil emulsions.

In certain such embodiments, the pharmaceutical composition comprises a suitable pharmaceutically acceptable diluent or carrier. In certain embodiments, a pharmaceutical composition comprises a sterile saline solution and one or more oligomeric compound. In certain embodiments, such pharmaceutical composition consists of a sterile saline solution and one or more oligomeric compound. In certain embodiments, the sterile saline is pharmaceutical grade saline. In certain embodiments, a pharmaceutical composition comprises sterile water. In certain embodiments, the sterile water is pharmaceutical grade water. In certain embodiments, a pharmaceutical composition comprises phosphate-buffered saline (PBS). In certain embodiments, the sterile PBS is pharmaceutical grade PBS. Compositions and methods for the formulation of pharmaceutical compositions are dependent upon a number of criteria, including, but not limited to, route of administration, extent of disease, or dose to be administered.

In certain embodiments, the pharmaceutical composition can include one or more guide(s), a guided nucleic acid binding agent or a nucleotide sequence encoding the guided nucleic acid binding agent. In some embodiments, the pharmaceutical composition further comprises a polynucleotide to be inserted (e.g., a donor template) to affect a desired sequence modification. In certain embodiments, guide compositions are generally formulated to achieve a physiologically compatible pH, and range from a pH of about 3 to a pH of

about 11, about pH 3 to about pH 7, depending on the formulation and route of administration. In some embodiments, the pH is adjusted to a range from about pH 5.0 to about pH 8.

In certain embodiments, pharmaceutical compositions disclosed herein are prepared for oral administration. In certain embodiments, pharmaceutical compositions are prepared for buccal administration. In certain embodiments, a pharmaceutical composition is prepared for administration by injection (*e.g.*, intravenous, subcutaneous, intramuscular, intrathecal (IT), intracerebroventricular (ICV), etc.). In certain of such embodiments, a pharmaceutical composition comprises a carrier and is formulated in aqueous solution, such as water or physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. In certain embodiments, other ingredients are included (*e.g.*, ingredients that aid in solubility or serve as preservatives). In certain embodiments, injectable suspensions are prepared using appropriate liquid carriers, suspending agents and the like. Certain pharmaceutical compositions for injection are presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers. Certain pharmaceutical compositions for injection are suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

In certain embodiments, the composition comprises one or more antioxidants. In certain embodiments, the one or more antioxidants function to reduce a degradation of the LNPs or components thereof, such as the ionizable lipids, the cargo, or both. In certain embodiments, the antioxidant is a chelating agent such as EDTA, citrate, vitamin E, or a polyphenol.

In certain embodiments, the composition comprises a ribonucleoprotein (RNP). In certain embodiments, the RNP comprises a guided nucleic acid binding agent and a guide disclosed herein. In certain embodiments, the RNP comprises an editing system disclosed herein.

In certain embodiments, the composition comprises a complex comprising an editing system disclosed herein in contact with a target DNA, wherein the complementary strand of the DNA comprises a sequence complementary to the guide adjacent to a sequence complementary to a sequence selected from Table 6. In certain embodiments, the complex comprises an RNP.

Methods of Use

In certain embodiments, there are provided methods of using the compositions, guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs, viral vectors and cells disclosed herein. Such methods include methods of editing a target nucleic acid, methods of creating a discontinuity in a target nucleic acid, methods of creating a double stranded break or a nick in a target DNA, methods of gene silencing, methods of gene activation, methods of homologous gene repair, methods of gene visualization, methods of epigenetic modulation of gene expression, methods of treating a disease or disorder

in a subject, methods of autologous or allogeneic cell therapy, methods of diagnosing a disease or disorder in a subject, and methods of assessing responsiveness to a treatment for a disease or disorder in a subject.

In certain embodiments, there is provided a method of editing a target nucleic acid, comprising (1) administering an LNP, a viral vector, or a composition as herein disclosed to a subject, or (2) contacting a cell with an LNP, a viral vector, or a composition as herein disclosed. In certain embodiments, the administering or contacting results in administration of an editing system comprising a guide and a Cas protein, or administration of an expression system comprising a guide and an exogenous mRNA encoding a Cas protein. In certain embodiments, such administration results in an RNP binding to a target nucleic acid, wherein binding of the RNP enables introduction of point mutations into a target nucleic acid. In certain such embodiments, the Cas protein is a fusion protein that includes a nucleobase deaminase. In certain embodiments, the nucleobase deaminase is a cytosine base editor that can introduce C>T or T>C transitions. In certain embodiments, the nucleobase deaminase is an adenine base editor that can introduce A>G or G>A transitions. In certain such other embodiments, the Cas protein is a Cas nickase fusion protein that includes a reverse transcriptase. In certain embodiments, the methods provide for a correction or compensation of a mutation in a cell, thereby creating an edited cell such that expression of a functional gene product can occur. In other embodiments, the methods provide for reducing or eliminating expression of a gene product by a knock-down or knock-out of the gene.

In certain embodiments, there is provided a method of creating a discontinuity in a target nucleic acid, comprising (1) administering an LNP, a viral vector, or a composition as herein disclosed to a subject, or (2) contacting a cell with an LNP, a viral vector, or a composition as herein disclosed. In certain embodiments, the administering or contacting results in administration of an editing system comprising a guide and a Cas protein, or administration of an expression system comprising a guide and an exogenous mRNA encoding a Cas protein. In certain embodiments, such administration results in an RNP binding to a target nucleic acid, wherein binding of the RNP creates a break in the target nucleic acid, such as a double strand break or a single strand break (e.g., a nick). In certain such embodiments, the break can result in random insertions or deletions (indels), or a substitution, duplication, frame-shift, or inversion of one or more nucleotides in those regions by non-homologous DNA end joining (NHEJ) repair mechanisms. In other such embodiments, homology-directed repair (HDR), homology-independent targeted integration (HITI), micro-homology mediated end joining (MMEJ), single strand annealing (SSA) or base excision repair (BER) may result in modification of the target nucleic acid sequence. In some embodiments, a modification comprises introducing an in-frame mutation in the target nucleic acid. In some embodiments, a modification comprises introducing a frame-shifting mutation in the target nucleic acid. In some embodiments, a modification comprises introducing a premature stop codon in the coding sequence in the target nucleic acid. In certain embodiments, repair of a break can result in, for example, a gene knockout, a gene knock-in or a gene knock-down. As a result of a gene knock-down by the foregoing modifications, the protein activity or function may

be attenuated or the protein levels may be reduced or eliminated. In some embodiments, the modification results in 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%, or at least about 90% or more reduced expression of a gene product in comparison to cells in which the gene has not been modified. In other embodiments, mutations in a gene may be corrected, wherein a corrective sequence is knocked-in by introducing mutations at select locations by design of the targeting sequence such that a wild-type or functional gene product is expressed.

In certain embodiments, there is provided a method of gene silencing, comprising (1) administering an LNP, a viral vector, or a composition as herein disclosed to a subject, or (2) contacting a cell with an LNP, a viral vector, or a composition as herein disclosed. In certain embodiments, the administering or contacting results in administration of an editing system comprising a guide and a dead Cas protein, or administration of an expression system comprising a guide and an exogenous mRNA encoding a dead Cas protein. In certain embodiments, such administration results in an RNP binding to a target nucleic acid. In embodiments where the dead Cas protein is a fusion protein that includes a transcriptional repressor domain, binding of the RNP to the target nucleic acid can prevent transcription or translation, thereby silencing expression of a target nucleic acid.

In certain embodiments, there is provided a method of gene activation, comprising (1) administering an LNP, a viral vector, or a composition as herein disclosed to a subject, or (2) contacting a cell with an LNP, a viral vector, or a composition as herein disclosed. In certain embodiments, the administering or contacting results in administration of an editing system comprising a guide and a dead Cas protein, or administration of an expression system comprising a guide and an exogenous mRNA encoding a dead Cas protein. In certain embodiments, such administration results in an RNP binding to a target nucleic acid. In embodiments where the dead Cas protein is a fusion protein that includes a transcriptional activator, binding of the RNP to the target nucleic acid activates expression of or over-expression of the target nucleic acid.

In certain embodiments, there is provided a cell, organ or tissue that has been edited or modified by the methods disclosed herein. In certain embodiments, the cell, organ or tissue may be *in vitro*. In certain embodiments, the cell may be *ex vivo*. In certain embodiments, the cell, organ or tissue may be used for autologous therapy or for allogeneic therapy.

In certain embodiments, there is provided a method of gene visualization, comprising (1) administering an LNP, a viral vector, or a composition as herein disclosed to a subject, or (2) contacting a cell with an LNP, a viral vector, or a composition as herein disclosed. In certain embodiments, the administering or contacting results in administration of an editing system comprising a guide and a dead Cas protein, or administration of an expression system comprising a guide and an exogenous mRNA encoding a dead Cas protein. In certain embodiments, such administration results in an RNP binding to a target nucleic acid. In

embodiments where the dead Cas protein is a fusion protein that includes a fluorescent protein or tag, binding of the RNP to the target nucleic acid enables visualization of the target nucleic acid.

In certain embodiments, visualization of a target nucleic acid provides for methods of diagnosing a disease or disorder in a subject, or methods of assessing responsiveness to a treatment for a disease or disorder in a subject. In certain embodiments, a disease or disorder may be diagnosed, or responsiveness to a treatment for a disease or disorder may be assessed, through means other than visualization of a target nucleic acid. In such embodiments, a target nucleic acid cleavage pattern may be assessed with reference to a control, or analyte tests known in the art may be performed on a biological sample of a subject, both before and after administration of a composition, LNP or viral vector disclosed herein, wherein a temporal difference in analyte result indicates the presence or absence of a disease or disorder, or indicates responsiveness or a lack of responsiveness to a treatment for a disease or disorder.

In certain embodiments, there is provided a method of epigenetic modulation of gene expression, comprising (1) administering an LNP, a viral vector, or a composition as herein disclosed to a subject, or (2) contacting a cell with an LNP, a viral vector, or a composition as herein disclosed. In certain embodiments, the administering or contacting results in administration of an editing system comprising a guide and a dead Cas protein, or administration of an expression system comprising a guide and an exogenous mRNA encoding a dead Cas protein. In certain embodiments, such administration results in an RNP binding to a target nucleic acid. In embodiments where the dead Cas protein is a fusion protein that includes, for example, a histone acetyltransferase (HAT) or a histone deacetylase (HDAC), binding of the RNP to the target nucleic acid results in acetylation or deacetylation of histone proteins, respectively, thereby altering chromatin structure and transcriptional regulation, thereby modulating epigenetic gene expression.

In certain embodiments, the target nucleic acid is a target DNA. In certain embodiments, the non-complementary strand of the target DNA comprises a sequence selected from Table 6 within 18-20 nucleobases from the cut site in the non-complementary strand.

In certain embodiments, the editing system or the expression system is administered using a composition or an LNP disclosed herein.

In certain embodiments, the methods disclosed herein comprise using a composition, guide, nucleic acid, or expression system disclosed herein, comprising a viral vector.

In certain embodiments, the methods disclosed herein comprise administering a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein to a subject.

In certain embodiments, the methods disclosed herein comprise contacting a cell with a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein. In certain embodiments, the cell is in a subject. In certain embodiments, the cell is *in vitro*. In certain embodiments, the cell is *ex vivo*.

In certain embodiments, there is provided a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein for use in therapy.

In certain embodiments, there is provided a method of treating a disease or disorder in a subject, wherein the method comprises administering an LNP, a viral vector or a composition disclosed herein. In certain embodiments, the method comprises editing or modifying a target nucleic acid, optionally wherein the editing or modifying occurs in a subject having a mutation in an allele of a gene, wherein the mutation causes a disease or disorder in the subject. In certain embodiments, the editing or modifying changes the mutation to a wild type allele of the gene or knocks down or knocks out an allele of a gene causing the disease or disorder in the subject.

In certain such embodiments, the editing or modifying comprises introducing a single-stranded break in the target nucleic acid of a cell in a subject. In other embodiments, the editing or modifying comprises introducing a double-stranded break in the target nucleic acid of a cell of a subject. In some embodiments, the editing or modifying introduces one or more mutations in the target nucleic acid, such as an insertion, deletion, substitution, duplication, or inversion of one or more nucleotides in an allele of a gene, wherein expression of the gene product in the edited or modified cells of the subject is reduced by 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%, or at least about 90% or more in comparison to a cell that has not been edited or modified. In certain embodiments, a gene of the edited or modified cells of the subject is edited or modified such that least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% of the edited or modified cells do not express a detectable level of the gene product.

In certain such embodiments, the administering knocks down or knocks out expression of a gene product, leading to prevention or amelioration of a disease or disorder in the subject such that an improvement is observed in the subject, notwithstanding that the subject may or may not still be afflicted with the disease or disease. In certain embodiments, the administering corrects or compensates for a mutation in a gene product, leading to the prevention or amelioration of a disease or disorder in the subject such that an improvement is observed in the subject, notwithstanding that the subject may or may not still be afflicted with the disease or disease. In such embodiments, the gene can be modified by the NHEJ host repair mechanisms, or utilized in conjunction with a donor template that is inserted by HDR or HITI mechanisms to either excise, correct, or compensate for the mutation in the cells of the subject, such that expression of a wild-type or functional gene product in modified cells is increased by at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or at least about 95% in comparison to a cell that has not been modified. In some embodiments, the administering leads to an improvement in at least one clinically-relevant parameter for a disease.

In certain embodiments, there is provided a method of autologous cell therapy, wherein the method comprises providing an autologous cell, contacting an LNP, a viral vector or a composition disclosed herein with the cell, and administering the cell to a subject.

In certain embodiments, there is provided a method of allogeneic cell therapy, wherein the method comprises providing an allogeneic cell, contacting an LNP, a viral vector or a composition disclosed herein with the cell, and administering the cell to a subject.

In certain embodiments, there is provided an autologous or allogeneic cell, organ or tissue that has been edited or modified by the methods disclosed herein. In certain embodiments, the cell, organ or tissue may be *in vitro*. In certain embodiments, the cell may be *ex vivo*. In certain embodiments, the cell, organ or tissue may be used for autologous therapy or for allogeneic therapy.

In certain embodiments, there is provided use of an LNP, a viral vector, or a composition disclosed herein, in the manufacture of a medicament for treating a disease or disorder in a subject.

In certain embodiments, there is provided use of an autologous or allogeneic cell, organ or tissue that has been edited or modified by the methods disclosed herein, in the manufacture of a medicament for treating a disease or disorder in a cell, organ or tissue of a subject.

In certain embodiments, the compositions, guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs or viral vectors disclosed herein can be delivered via transfection such as calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, electrical nuclear transport, chemical transduction, electrotransduction, Lipofectamine-mediated transfection, Effectene-mediated transfection, LNP-mediated transfection, or any combination thereof. In some embodiments, the composition is introduced to cells via lipid-mediated transfection using an LNP.

In some embodiments, the pharmaceutical composition can be administered by aerosol delivery, nasal delivery, vaginal delivery, rectal delivery, buccal delivery, ocular delivery, local delivery, topical delivery, intracisternal delivery, intraperitoneal delivery, oral delivery, intramuscular injection, intravenous injection, subcutaneous injection, intranodal injection, intratumoral injection, intraperitoneal injection, and/or intradermal injection, or any combination thereof. The administration can be local or systemic. The systemic administration includes enteral and parenteral administration.

In certain embodiments, a subject is administered a pharmaceutical composition two or more times. In certain embodiments, multiple administrations of the pharmaceutical composition can be separated by a suitable period of time, such as one week, two weeks, three weeks, four weeks, five weeks, six weeks, seven weeks, eight weeks, three months, four months, five months, six months, a year, eighteen months, two years, three years, five years, ten years, fifteen years, or more.

In some embodiments, the two or more administrations are about two weeks to about two months apart. The suitable time period between some administrations can be the same as or different from the suitable time period between other two administrations. In some embodiments, the method described herein comprises administration of a single dose of a pharmaceutical composition to the subject in a provided period

of time, for example, one year, two years, three years, five years, six years, eight years, ten years, fifteen years, twenty years, or longer. In some embodiments, the method described herein comprises administration of a single dose of a pharmaceutical composition to a subject in the subject's life time. In some embodiments, the method described herein comprises administration of a single dose of a pharmaceutical composition to the subject.

In certain embodiments, the pharmaceutical composition can be administered to a subject at a pharmaceutically effective amount. In some embodiments, the pharmaceutical composition is administered to the subject at a dose of about 0.01-5 mg/kg, for example 0.05-2 mg/kg, 0.5-3 mg/kg or 0.1-1 mg/kg, guide per administration. In some embodiments, the pharmaceutical composition is administered to the subject at a dose of about 0.01-5 mg/kg, for example 0.05-2 mg/kg, 0.5-3 mg/kg or 0.1-1 mg/kg, total nucleic acid (i.e., the total of the guide and mRNA encoding the guided nucleic acid binding agent) per administration.

In certain embodiments, the methods disclosed herein are *in vitro*, inside of a cell, such as in a cell culture system. In some embodiments, methods disclosed herein are *in vivo* inside of a cell of a subject, for example in a cell in an animal. In some embodiments, the cell is a eukaryotic cell. Exemplary eukaryotic cells may include cells selected from the group consisting of a mouse cell, a rat cell, a pig cell, a dog cell, and a non-human primate cell. In some embodiments, the cell is a human cell. Non-limiting examples of cells include an embryonic stem cell, an induced pluripotent stem cell, a germ cell, a fibroblast, an oligodendrocyte, a glial cell, a hematopoietic stem cell, a neuron progenitor cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell, a retinal cell, a cancer cell, a T-cell, a B-cell, an NK cell, a fetal cardiomyocyte, a myofibroblast, a mesenchymal stem cell, an autotransplanted expanded cardiomyocyte, an adipocyte, a totipotent cell, a pluripotent cell, a blood stem cell, a myoblast, an adult stem cell, a bone marrow cell, a mesenchymal cell, a parenchymal cell, an epithelial cell, an endothelial cell, a mesothelial cell, fibroblasts, osteoblasts, chondrocytes, exogenous cell, endogenous cell, stem cell, hematopoietic stem cell, bone-marrow derived progenitor cell, myocardial cell, skeletal cell, fetal cell, undifferentiated cell, multi-potent progenitor cell, unipotent progenitor cell, a monocyte, a cardiac myoblast, a skeletal myoblast, a macrophage, a capillary endothelial cell, a xenogenic cell, an allogenic cell, or a post-natal stem cell. In alternative embodiments, the cell is a prokaryotic cell.

In certain embodiments, wherein the methods disclosed herein are *in vitro*, the compositions, guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs or viral vectors disclosed herein may be administered to *in vitro* cells, such as autologous cells or allogeneic cells, for about 30 minutes to about 24 hours, or at least about 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 12 hours, 16 hours, 18 hours, 20 hours, or any other period from about 30 minutes to about 24 hours, which may be repeated with a frequency of about every day to about every 4 days, *e.g.*, every 1.5 days, every 2 days, every 3 days, or any other frequency from about every day to

about every four days. The autologous or allogeneic cells may then be provided to the subject for treatment of a disease or disorder.

In certain embodiments, the compositions, guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs, viral vectors or cells disclosed herein provide surprising and advantageous cargo delivery efficiency. In certain embodiments, target DNA cleavage is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%. In certain embodiments, the target DNA is in a liver cell, for example a hepatocyte. In certain embodiments, the subject is a mammal such as a human.

In certain embodiments, the compositions, guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs, viral vectors or cells disclosed herein provide surprising and advantageous reductions in immunogenicity, thereby reducing blood clearance, reducing dose-limiting toxicity and reducing complement activation-related pseudoallergy (CARPA).

Kits and Devices

In certain embodiments, there are provided kits and devices comprising the compositions, guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs, viral vectors or cells disclosed herein. Optionally, a kit is for, or when used for, treating a disease or disorder in a subject.

Kits provided herein facilitate employment of the methods and uses of the invention. Typically, kits for carrying out a method or use of the invention contain all the necessary reagents and means to carry out the method. In one embodiment, the kit may comprise a composition of the present invention and, optionally, means to administer the composition such as devices for point of care methods.

In certain embodiments, kits comprise one or more containers. A compartmentalized kit includes any kit in which compositions are contained in separate containers, and may include small glass containers, plastic containers or strips of plastic or paper. Such containers may allow the efficient transfer of compositions from one compartment to another compartment whilst avoiding cross-contamination of compositions, and the addition of agents or solutions of each container from one compartment to another in a quantitative fashion.

In certain embodiments, kits include instructions for using the kit to perform the appropriate methods and uses.

In certain embodiments, kits comprise a saline, a buffered solution and a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein. In one embodiment, the buffered solution may include sodium chloride, calcium chloride, phosphate and/or EDTA. In another embodiment, the buffer solution may include, but is not limited to, saline, saline with 2mM calcium, 5% sucrose, 5% sucrose with 2mM calcium, 5% Mannitol, 5% Mannitol with 2mM

calcium, Ringer's lactate, sodium chloride, sodium chloride with 2mM calcium and mannose (See e.g., U.S. Pub. No. 2012/0258046; herein incorporated by reference in its entirety).

In one embodiment, the buffered solution may be precipitated or it may be lyophilized. The amount of each component in the kit may be varied to enable consistent, reproducible higher concentration saline or simple buffered formulations.

In certain embodiments, the kits comprise one or more additional reagents selected from a buffer, a buffer for introducing a polypeptide or nucleic acid into a cell, a wash buffer, a control reagent, a control vector, a control RNA, a reagent for *in vitro* production of the polypeptide from DNA, adaptors for sequencing and the like. A buffer can be a stabilization buffer, a reconstituting buffer, a diluting buffer, or the like. The kits can also comprise one or more components that can be used to facilitate or enhance on-target binding or cleavage of a target nucleic acid by an endonuclease, or improve specificity of targeting.

The components may also be varied in order to increase the stability of the composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein in the buffered solution over a period of time and/or under a variety of conditions.

In certain embodiments, the kits can further include instructions for using the components of the kit to practice the methods described herein. The instructions for practicing the methods are generally recorded on a suitable recording medium. For example, the instructions can be printed on a substrate, such as paper or plastic, etc. The instructions can be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging), etc. The instructions can be present as an electronic storage data file present on a suitable computer readable storage medium, e.g., CD-ROM, diskette, flash drive, etc. In some instances, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source (e.g., via the Internet), can be provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions can be recorded on a suitable substrate.

Devices provided herein facilitate employment of the methods and uses of the invention. In certain embodiments, devices may incorporate a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP, viral vector or cell disclosed herein. In certain embodiments, devices contain in a stable formulation the reagents required to formulate a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein to be immediately delivered to a subject in need thereof.

Devices for administration may be employed to deliver a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein according to single, multi- or split-dosing regimens. See, for example, PCT/US2013/30062, the contents of which are incorporated herein by reference in their entirety.

Devices known in the art for single or multi-administration to cells, organs and tissues can be used in conjunction with the methods disclosed herein. Examples of such devices include syringes, needles, devices having multiple needles, hybrid devices employing lumens or catheters, and devices utilizing heat, electric current or radiation driven mechanisms.

Multi-administration devices may be utilized to deliver single, multi- or split doses. See, for example, PCT/US2013/30062, the contents of which are incorporated herein by reference in their entirety.

Devices using catheters and lumens may be employed to administer a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein on a single, multi- or split dosing schedule. Such methods and devices are described in PCT/US2013/30062, the contents of which are incorporated herein by reference in their entirety.

Devices utilizing electric current may be employed to deliver a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP or viral vector disclosed herein using single, multi- or split dosing regimens. Such methods and devices are described in PCT/US2013/30062, the contents of which are incorporated herein by reference in their entirety.

Also provided herein is an article of manufacture comprising a composition, guided nucleic acid binding agent, guide, editing system, nucleic acid, expression system, LNP, viral vector or cell disclosed herein.

Nonlimiting Disclosure and Incorporation by Reference

Each of the literature and patent publications listed herein is incorporated by reference in its entirety.

While certain compounds, compositions, and methods have been described herein with specificity in accordance with certain embodiments, the following examples serve only to illustrate the compounds described herein and are not intended to limit the same. Each of the references, GenBank accession numbers, ENSEMBL identifiers, and the like recited in the present application is incorporated herein by reference in its entirety.

The sequence listing accompanying this filing identifies each nucleic acid sequence as either "RNA" or "DNA" as required; however, one of skill in the art will readily appreciate that designation of "RNA" or "DNA" to describe modified oligonucleotides is, in certain instances, arbitrary. For example, an oligonucleotide comprising a nucleoside comprising a 2'-OH sugar moiety and a thymine base could be described as a DNA having a modified sugar (i.e., 2'-OH in place of one 2'-H of DNA) or as an RNA having a modified base (i.e., thymine (5-methyl uracil) in place of an uracil of RNA); and certain nucleic acid compounds described herein comprise one or more nucleosides comprising modified sugar moieties having 2'-substituent(s) that are neither OH nor H. One of skill in the art will readily appreciate that labeling such nucleic acid compounds "RNA" or "DNA" does not alter or limit the description of such nucleic acid compounds.

Herein, the description of compounds as having "the nucleobase sequence of" a SEQ ID NO. describes only the nucleobase sequence. Accordingly, absent additional description, such description of compounds by reference to a nucleobase sequence of a SEQ ID NO. does not limit sugar or internucleoside linkage modifications or presence or absence of additional substituents such as a conjugate group. Further, absent additional description, the nucleobases of a compound "having the nucleobase sequence of" a SEQ ID NO. include such compounds having modified forms of the identified nucleobases as described herein.

Herein, sugar, internucleoside linkage, and nucleobase modifications may be indicated within a nucleotide or nucleobase sequence or may be indicated in text accompanying a sequence (e.g., in separate text that appears within or above or below a table of compounds).

While effort has been made to accurately describe compounds in the accompanying sequence listing, should there be any discrepancies between a description in this specification and in the accompanying sequence listing, the description in the sequence listing and not the specification is the accurate description.

Certain compounds described herein (*e.g.*, nucleic acids such as exogenous mRNA, antisense agents, oligomeric compounds, modified oligonucleotides and LNPs) may have one or more asymmetric centers and thus give rise to enantiomers, diastereomers, and other stereoisomeric configurations that may be defined, in terms of absolute stereochemistry, as (*R*) or (*S*), as α or β such as for sugar anomers, or as (D) or (L), such as for amino acids, etc. Compounds provided herein that are drawn or described as having certain stereoisomeric configurations are enriched (e.g., to at least about 50% ee) for the indicated stereochemistry. Compounds provided herein that are drawn or described with undefined stereochemistry include racemic, stereorandom, and optically enriched forms. The present disclosure includes all such possible isomers, as well as their racemic and optically pure forms whether or not they are specifically depicted herein. Optically active (+) and (-), (*R*)- and (*S*)-, or (D)- and (L)- isomers can be prepared using chiral synthons or chiral reagents, or resolved using conventional techniques, for example, chromatography and fractional crystallization. Conventional techniques for the preparation/isolation of individual enantiomers include chiral synthesis from a suitable optically pure precursor or resolution of the racemate (or the racemate of a salt or derivative) using, for example, chiral high pressure liquid chromatography (HPLC). When the compounds described herein contain olefinic double bonds or other centers of geometric asymmetry, and unless specified otherwise, it is intended that the compounds include both E and Z geometric isomers. All tautomeric forms of the compounds provided herein are included unless otherwise indicated.

The compositions, guided nucleic acid binding agents, guides, editing systems, nucleic acids, expression systems, LNPs, viral vectors and cells described herein include variations in which one or more atoms are replaced with a non-radioactive isotope or radioactive isotope of the indicated element. For example, compounds herein that comprise hydrogen atoms encompass all possible deuterium substitutions for each of the ^{1}H hydrogen atoms. Isotopic substitutions encompassed by the compounds herein include but are

not limited to: ^{2}H or ^{3}H in place of ^{1}H, ^{13}C or ^{14}C in place of ^{12}C, ^{15}N in place of ^{14}N, ^{17}O or ^{18}O in place of ^{16}O, and ^{33}S, ^{34}S, ^{35}S, or ^{36}S in place of ^{32}S. In certain embodiments, non-radioactive isotopic substitutions may impart new properties on the oligomeric compound that are beneficial for use as a therapeutic or research tool. In certain embodiments, radioactive isotopic substitutions may make the compound suitable for research or diagnostic purposes such as imaging.

EXAMPLES

Example 1: Computational prediction of Cas proteins

Metagenomes were assembled using publicly available datasets. CRISPR cassettes containing CRISPR array and nearby proteins were predicted by CRISPRFinder and CRISPRone and were classified into sub-classes and -types. For Class 2 Type V CRISPR systems, protein domains were identified via Hidden Markov Models and sequences were aligned to Cas12a of *Acidaminococcus sp.* (AsCas12a). Multiple criteria were used to narrow down potentially functional Cas proteins. The conserved catalytic residues in the RuvC domain were present in each novel Cas protein via multiple sequence alignment analysis.

After these analyses, 92 putative Cas proteins were identified. The corresponding Cas protein sequences (SEQ ID NO: 4-95) and GeneArt (ThermoFisher) human codon optimized nucleic acid sequences (SEQ ID NO: 96-187) are provided in the sequence listing, as indicated in the table below. The Cas protein DNA sequences were cloned into a pTwistCMVpuro vector by Twist Biosciences. To facilitate enzyme localization to the nucleus, a nucleoplasmin-Nuclear Localization Signal (NLS) and two repeats of c-Myc NLS were added to the C-terminus of each Cas protein sequence. In addition, a HiBiT (Promega) tag was added between the nucleoplasmin NLS sequence and the first c-Myc NLS sequence for expression detection. These additional elements were linked to the proteins and translated to codon-optimized DNA, as indicated in the DNA sequences represented by SEQ ID NOs: 188-279 below.

Table 3

Cas Protein sequences and AsCas12a

Cas Protein ID	Cas Protein SEQ ID NO.	Cas protein codon-optimized DNA SEQ ID NO.	Cas protein + NLS + HiBit codon-optimized DNA SEQ ID NO.
AsCas12a	1	2	3
IONCAS-V001	4	96	188
IONCAS-V002	5	97	189
IONCAS-V003	6	98	190
IONCAS-V004	7	99	191
IONCAS-V005	8	100	192
IONCAS-V006	9	101	193
IONCAS-V007	10	102	194

IONCAS-V008	11	103	195
IONCAS-V009	12	104	196
IONCAS-V010	13	105	197
IONCAS-V011	14	106	198
IONCAS-V012	15	107	199
IONCAS-V013	16	108	200
IONCAS-V014	17	109	201
IONCAS-V015	18	110	202
IONCAS-V016	19	111	203
IONCAS-V017	20	112	204
IONCAS-V018	21	113	205
IONCAS-V019	22	114	206
IONCAS-V020	23	115	207
IONCAS-V021	24	116	208
IONCAS-V022	25	117	209
IONCAS-V023	26	118	210
IONCAS-V024	27	119	211
IONCAS-V025	28	120	212
IONCAS-V026	29	121	213
IONCAS-V027	30	122	214
IONCAS-V028	31	123	215
IONCAS-V029	32	124	216
IONCAS-V030	33	125	217
IONCAS-V031	34	126	218
IONCAS-V032	35	127	219
IONCAS-V033	36	128	220
IONCAS-V034	37	129	221
IONCAS-V035	38	130	222
IONCAS-V036	39	131	223
IONCAS-V037	40	132	224
IONCAS-V038	41	133	225
IONCAS-V039	42	134	226
IONCAS-V040	43	135	227
IONCAS-V041	44	136	228
IONCAS-V042	45	137	229
IONCAS-V043	46	138	230
IONCAS-V044	47	139	231
IONCAS-V045	48	140	232
IONCAS-V046	49	141	233
IONCAS-V047	50	142	234
IONCAS-V048	51	143	235
IONCAS-V049	52	144	236
IONCAS-V050	53	145	237
IONCAS-V051	54	146	238
IONCAS-V052	55	147	239

IONCAS-V053	56	148	240
IONCAS-V054	57	149	241
IONCAS-V055	58	150	242
IONCAS-V056	59	151	243
IONCAS-V057	60	152	244
IONCAS-V058	61	153	245
IONCAS-V059	62	154	246
IONCAS-V060	63	155	247
IONCAS-V061	64	156	248
IONCAS-V062	65	157	249
IONCAS-V063	66	158	250
IONCAS-V064	67	159	251
IONCAS-V065	68	160	252
IONCAS-V066	69	161	253
IONCAS-V067	70	162	254
IONCAS-V068	71	163	255
IONCAS-V069	72	164	256
IONCAS-V070	73	165	257
IONCAS-V071	74	166	258
IONCAS-V072	75	167	259
IONCAS-V073	76	168	260
IONCAS-V074	77	169	261
IONCAS-V075	78	170	262
IONCAS-V076	79	171	263
IONCAS-V077	80	172	264
IONCAS-V078	81	173	265
IONCAS-V079	82	174	266
IONCAS-V081	83	175	267
IONCAS-V082	84	176	268
IONCAS-V083	85	177	269
IONCAS-V084	86	178	270
IONCAS-V085	87	179	271
IONCAS-V086	88	180	272
IONCAS-V087	89	181	273
IONCAS-V088	90	182	274
IONCAS-V089	91	183	275
IONCAS-V090	92	184	276
IONCAS-V091	93	185	277
IONCAS-V092	94	186	278
IONCAS-V093	95	187	279

Example 2: Evaluation of Cas protein activity in HEK293T nuclear extract

Cas proteins described above were screened for enzymatic activity in an *in vitro* cleavage assay wherein an artificial synthesized DNA target and synthesized guide were incubated with the nuclear extract of HEK293T cells that had been transfected with pTwistCMV-puro vectors expressing Cas proteins. The cleavage of the synthesized DNA target was then measured as an indicator of Cas protein activity.

Sequences of protein-specific protein-recognition region sequences for the Cas proteins were computationally predicted. The protein-recognition region includes a direct repeat region and a 5'-extension. The direct repeat sequences are shown in the table below, with the Cas Proteins they were designed for indicated in the column titled "Cas Protein ID". The protein-recognition region sequences can be found in the sequence listing as DNA sequences (SEQ ID NO: 280-371) or as RNA sequences as indicated in the table below under the column "Protein-recognition region SEQ ID NO.".

Table 4

Sequences of direct repeat regions for each Cas protein, guide, and protein-recognition region

Cas Protein ID	Cas Protein SEQ ID NO.	Direct repeat region ID	Direct Repeat Region Sequence	Direct Repeat SEQ ID NO.	Protein-recognition Region SEQ ID NO.
IONCAS-V001	4	V1_DR	GAAUUUCUACUUUCGUAGAU	1176	1061
IONCAS-V002	5	V2_DR	AAUUUCUACUAUUUGUAGAU	1177	1062
IONCAS-V003	6	V3_DR	UAAUUUCUACUAUUGUAGAU	1178	1063
IONCAS-V004	7	V4_DR	AAAUUUCUACUGUUGUAGAU	1179	1064
IONCAS-V005	8	V5_DR	AAAUUUCUACUAUUGUAGAU	1180	1065
IONCAS-V006	9	V6_DR	UCAAUUUCUACUUUUGUAGA	1181	1066
IONCAS-V007	10	V7_DR	AAAUUUCUACUGUUGUAGAU	1179	1067
IONCAS-V008	11	V8_DR	UAAUUUCUACUUCGGUAGAU	1182	1068
IONCAS-V009	12	V9_DR	AAGAUUUCUACUGUUGUAGA	1183	1069
IONCAS-V010	13	V10_DR	AAUUUCUACUUAGUGUAGAU	1184	1070
IONCAS-V011	14	V11_DR	UAAUUUCUACUGUUGUAGAU	1185	1071
IONCAS-V012	15	V12_DR	CAAUUUCUACUAAUGUAGAU	1186	1072
IONCAS-V013	16	V13_DR	AAAUUUCUACUGUUGUAGAU	1179	1073
IONCAS-V014	17	V14_DR	AUUAAUGCUGGUCUCUAGAG	1187	1074
IONCAS-V015	18	V15_DR	AUUAUAAGAGGUUAUUAAAC	1188	1075
IONCAS-V016	19	V16_DR	UAAUUUCUACUAUUGUAGAU	1178	1076
IONCAS-V017	20	V17_DR	AAAUUUCUACUAUUGUAGAU	1180	1077
IONCAS-V018	21	V18_DR	GAAUUUCUACUGUUGUAGAU	1189	1078
IONCAS-V019	22	V19_DR	UAAUUUCUACUGUUGUAGAU	1185	1079
IONCAS-V020	23	V20_DR	UAAUUUCUACUAUUGUAGAU	1178	1080

IONCAS-V021	24	V21_DR	UAAUUUCUACUAUUGUAGAU	1178	1081
IONCAS-V022	25	V22_DR	AAUUUCUACUAAGUGUAGAU	1190	1082
IONCAS-V023	26	V23_DR	UAAUUUCUACUUUUGUAGAU	1191	1083
IONCAS-V024	27	V24_DR	AAAUUUCUACUAUUGUAGAU	1180	1084
IONCAS-V025	28	V25_DR	UAAUUUCUACUAUUGUAGAU	1178	1085
IONCAS-V026	29	V26_DR	AAAUUUCUACUAUUGUAGAU	1180	1086
IONCAS-V027	30	V27_DR	AAAAUUUCUACUAUUGUAGA	1192	1087
IONCAS-V028	31	V28_DR	GAAUUUCUACUAUUGUAGAU	1193	1088
IONCAS-V029	32	V29_DR	UAAUUUCUACUAUUGUAGAU	1178	1089
IONCAS-V030	33	V30_DR	UAAUUUCUACUAUUGUAGAU	1178	1090
IONCAS-V031	34	V31_DR	AUUUAAUGAAAAGUAUAAAC	1194	1091
IONCAS-V032	35	V32_DR	UAAUUUCUACUAUUGUAGAU	1178	1092
IONCAS-V033	36	V33_DR	AAAUUUCUACUGUUGUAGAU	1179	1093
IONCAS-V034	37	V34_DR	AAAUUUCUACUAUUGUAGAU	1180	1094
IONCAS-V035	38	V35_DR	AAAUGCCAAACUUUGAGGAA	1195	1095
IONCAS-V036	39	V36_DR	GAAUUUCUACUGUUGUAGAU	1189	1096
IONCAS-V037	40	V37_DR	AAAUUUCUACUAUUGUAGAU	1180	1097
IONCAS-V038	41	V38_DR	UAAAUUUCUACUAUUGUAGA	1196	1098
IONCAS-V039	42	V39_DR	AAAUUUCUACUGUUGUAGAU	1179	1099
IONCAS-V040	43	V40_DR	AAAUUUCUACUGUUGUAGAU	1179	1100
IONCAS-V041	44	V41_DR	AAUUUCUACUAUGUGUAGAU	1197	1101
IONCAS-V042	45	V42_DR	UAAUUUCUACUUUUGUAGAU	1191	1102
IONCAS-V043	46	V43_DR	UAAAUUUCUAUUAUUAUAGA	1198	1103
IONCAS-V044	47	V44_DR	UAAUUUCUACUAUUGUAGAU	1178	1104
IONCAS-V045	48	V45_DR	GCAUUCGGAAGAGGUAAAGA	1199	1105
IONCAS-V046	49	V46_DR	UAAUUUCUACUAUUGUAGAU	1178	1106
IONCAS-V047	50	V47_DR	UAAUUUCUACUGUUGUAGAU	1185	1107
IONCAS-V048	51	V48_DR	UAAUUUCUACUAUCGUAGAU	1200	1108
IONCAS-V049	52	V49_DR	UAAUUUCUACUAUUGUAGAU	1178	1109
IONCAS-V050	53	V50_DR	UAAUUUCUACUAUUGUAGAU	1178	1110
IONCAS-V051	54	V51_DR	UAAUUUCUACUAUUGUAGAU	1178	1111
IONCAS-V052	55	V52_DR	UAAUUUCUACUAUUGUAGAU	1178	1112
IONCAS-V053	56	V53_DR	AAUUUCUACUAUUGUAGAUA	1201	1113
IONCAS-V054	57	V54_DR	AUUCUAUAGGCUUAUUAGCC	1202	1114
IONCAS-V055	58	V55_DR	UAAUUUCUACUAUUGUAGAU	1178	1115
IONCAS-V056	59	V56_DR	AAUGGACUAUUUGUAUUUGA	1203	1116
IONCAS-V057	60	V57_DR	UAAUUUCUACUGUUGUAGAU	1185	1117
IONCAS-V058	61	V58_DR	UACUGUUGUAGAUAUAUAAA	1204	1118
IONCAS-V059	62	V59_DR	AUUAACUUAGGUUUAUAGCC	1205	1119
IONCAS-V060	63	V60_DR	AAUUUCUACUAUUAGUAGAU	1206	1120

IONCAS-V061	64	V61_DR	UAAAUUUCUACUGUUGUAGA	1207	1121
IONCAS-V062	65	V62_DR	AAUUUCUACUAUGUGUAGAU	1197	1122
IONCAS-V063	66	V63_DR	AUUUAAUAGGUCGCCUAGAC	1208	1123
IONCAS-V064	67	V64_DR	UAAUUUCUACUAUUGUAGAU	1178	1124
IONCAS-V065	68	V65_DR	CAAAUUUCUACUGUUGUAGA	1209	1125
IONCAS-V066	69	V66_DR	AAUUUCUACUGUUGUAGAUA	1210	1126
IONCAS-V067	70	V67_DR	AUUAUUUGUGGCUUAUAGCC	1211	1127
IONCAS-V068	71	V68_DR	GAAUUUCUACUAUUGUAGAU	1193	1128
IONCAS-V069	72	V69_DR	GCGAAGCGUAAAAAUCGGUC	1212	1129
IONCAS-V070	73	V70_DR	AUUCUCUAUAACGAUUUGGC	1213	1130
IONCAS-V071	74	V71_DR	AAAUUUCUACUGUUGUAGAU	1179	1131
IONCAS-V072	75	V72_DR	AUUUUGUAUUGGUCUUAAAC	1214	1132
IONCAS-V073	76	V73_DR	AAAAUUUCUACUUAUGUAGA	1215	1133
IONCAS-V074	77	V74_DR	AAAUUUCUACUAUUGUAGAU	1180	1134
IONCAS-V075	78	V75_DR	AUUGUUUUAAACGAUUAAAC	1216	1135
IONCAS-V076	79	V76_DR	AGAUUUCUACUUUUGUAGAU	1217	1136
IONCAS-V077	80	V77_DR	GAAUUUCUACUUUCGUAGAU	1176	1061
IONCAS-V078	81	V78_DR	AAAUUUCUACUGUCGUAGAU	1218	1137
IONCAS-V079	82	V79_DR	UAAAUUUCUACUCUCGUAGA	1219	1138
IONCAS-V081	83	V81_DR	UAAGAUUUCUACUGUUGUAG	1220	1139
IONCAS-V082	84	V82_DR	AAAUUUCUACUGUUGUAGAU	1179	1140
IONCAS-V083	85	V83_DR	AAAUUUCUACUUUUGUAGAU	1221	1141
IONCAS-V084	86	V84_DR	AAAUUUCUACUUUCGUAGAU	1222	1142
IONCAS-V085	87	V85_DR	AUCUAUAUACAAGUAUAGAC	1223	1143
IONCAS-V086	88	V86_DR	AAAUUUCUACUGUUGUAGAU	1179	1144
IONCAS-V087	89	V87_DR	AUUUAAUAUGAAGUUCAAAC	1224	1145
IONCAS-V088	90	V88_DR	AAAAUUUCUACUGUAUGUAG	1225	1146
IONCAS-V089	91	V89_DR	GAAUUUCUACUUUUGUAGAU	1226	1147
IONCAS-V090	92	V90_DR	AAUCUUAUAUAGGUAUUAGA	1227	1148
IONCAS-V091	93	V91_DR	UAAUUUCUACUAUUGUAGAU	1178	1149
IONCAS-V092	94	V92_DR	UAAUUUCUACUAUUGUAGAU	1178	1150
IONCAS-V093	95	V93_DR	AAAUUUCUACUAUUGUAGAU	1180	1151
AsCas12a	1	AsCas12a_DR	UAAUUUCUACUCUUGUAGAU	1228	1152

Guides comprising a Cas-specific protein-recognition region (at the 5` end) and a general target-recognition region (at the 3` end) were designed and ordered from IDT. The general target-recognition region has the sequence (from 5' to 3'): CUGAUGGUCCAUGUCUGUUACUC (SEQ ID NO: 497; DNA sequence SEQ ID NO: 466; previously disclosed in Zetsche et. al., *Cell* 2015; 163(3): 759-71). The guide sequence was

designed by appending the general target-recognition sequence to the 3'-end of the protein-recognition sequence. The guide was capped at the 3'-end with UUU. Guide sequences are shown in the table below.

Each guide in the tables below has the following modifications: the nucleosides at positions 1, 2 and 3 at the 5'-end are 2'-OMe sugar moieties wherein each nucleoside is linked to the next nucleoside with phosphorothioate internucleoside linkage; and the last three nucleosides at the 3'-end are 2'-OMe sugar moieties, wherein each nucleoside is linked to the preceding nucleoside by a phosphorothioate internucleoside linkage. In the table below, the target-recognition region sequence for the *in vitro* general target is indicated by bolded and underlined nucleobases, and the 2'-OMe modified nucleosides are indicated by italicized nucleobases.

The Cas protein that each guide was designed for is indicated in the column "Cas Protein ID." The protein-recognition region sequences, full guide sequences, and chemically-modified guide sequences are provided in the sequence listing as indicated in the table below under the columns "Protein-recognition region SEQ ID NO.", "Guide SEQ ID NO.", and "Mod. Guide SEQ ID NO." respectively.

Table 5

Sequences of guide protein-recognition regions predicted for the corresponding Cas proteins and AsCas12a

Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide Sequence (5' to 3')	Direct Repeat Region SEQ ID NO.	Protein-Recognition Region SEQ ID NO.	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.
IONCAS-V001	4	V1gRNA	*GGC*UAAAGCUCUUUAAGAAUUUCU ACUUUCGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1176	1061	373	819
IONCAS-V002	5	V2gRNA	*GUC*UAAUACUUAAUAUAAUUUCUA CUAUUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1177	1062	374	820
IONCAS-V003	6	V3gRNA	*GUC*UAUAAGACUAUUUUAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1178	1063	375	821
IONCAS-V004	7	V4gRNA	*AAU*AACAAUCCCUCUAAAUUUCUA CUGUUGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1179	1064	376	822
IONCAS-V005	8	V5gRNA	*GCU*AUAAAGCCAAUUAAAUUUCUA CUAUUGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1180	1065	377	823
IONCAS-V006	9	V6gRNA	*CUC*UAAUAGGCAUAUUCAAUUUCU ACUUUUGUAGA**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1181	1066	378	824

IONCAS-V007	10	V7gRNA	*GUC*UAAAGGCACCAAUAAAUUUCU ACUGUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1179	1067	379	825
IONCAS-V008	11	V8gRNA	*GUC*AAUAAGACUCAUUUAAUUUCU ACUUCGGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1182	1068	380	826
IONCAS-V009	12	V9gRNA	*CCC*GAAUACACCUAUAAGAUUUCU ACUGUUGUAGA**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1183	1069	381	827
IONCAS-V010	13	V10gRNA	*AUU*UAAUAAUCCUUUUAAUUUCUA CUUAGUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1184	1070	382	828
IONCAS-V011	14	V11gRNA	*GCC*AAGAACCUAUAGAUAAUUUCU ACUGUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1185	1071	383	829
IONCAS-V012	15	V12gRNA	*GUU*UAAUUAUUACUAUCAAUUUCU ACUAAUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1186	1072	384	830
IONCAS-V013	16	V13gRNA	*GUC*GAAUUGACCAUAUAAAUUUCU ACUGUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1179	1073	385	831
IONCAS-V014	17	V14gRNA	*AUC*UACAAGAGUAGACAUUAAUGC UGGUCUCUAGAG**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1187	1074	386	832
IONCAS-V015	18	V15gRNA	*AUC*UACAACAGUAGAAAUUAUAAG AGGUUAUUAAAC**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1188	1075	387	833
IONCAS-V016	19	V16gRNA	*GUC*UAAAAGACCUUUUUAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1178	1076	388	834
IONCAS-V017	20	V17gRNA	*GUU*UAAUACCUAUAUUAAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1180	1077	389	835
IONCAS-V018	21	V18gRNA	*CCU*UUGGAAGUACUAAGAAUUUCU ACUGUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1189	1078	390	836
IONCAS-V019	22	V19gRNA	*UGU*UUAAUAAUCCCUUAUAAUUUC UACUGUUGUAGAU**CUGAUGGUCC AUGUCUGUUACUC***uuu*	1185	1079	391	837
IONCAS-V020	23	V20gRNA	*GUU*UAAUACCUGUAUUUAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1178	1080	392	838
IONCAS-V021	24	V21gRNA	*GGU*UAAUAAACCCUUUUAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1178	1081	393	839
IONCAS-V022	25	V22gRNA	*GUU*AGAUACCUAUAUUAAUUUCUA CUAAGUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1190	1082	394	840
IONCAS-V023	26	V23gRNA	*AUC*UAAUAACCCUAUAUAAUUUCU ACUUUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1191	1083	395	841

IONCAS-V024	27	V24gRNA	*GCU*UAAUGCUUAUAUUAAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1180	1084	396	842
IONCAS-V025	28	V25gRNA	*GUC*UAAUACUUAAAUAUAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1178	1085	397	843
IONCAS-V026	29	V26gRNA	*GUU*UAAUACUUAUAUUAAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1180	1086	398	844
IONCAS-V027	30	V27gRNA	*GGU*UAAUACCUUAAGAAAAUUUCU ACUAUUGUAGA**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1192	1087	399	845
IONCAS-V028	31	V28gRNA	*UCU*AACAGAUAUAUUGAAUUUCUA CUAUUGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1193	1088	400	846
IONCAS-V029	32	V29gRNA	*GUC*UAAAAGACUGAUUUAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1178	1089	401	847
IONCAS-V030	33	V30gRNA	*UUU*AACAGUUCCUUAUAAUUUCUA CUAUUGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1178	1090	402	848
IONCAS-V031	34	V31gRNA	*AUC*UACAACAGUAGAAAUUUAAUG AAAAGUAUAAAC**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1194	1091	403	849
IONCAS-V032	35	V32gRNA	*GUC*UAAAAGACUUCUGUAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1178	1092	404	850
IONCAS-V033	36	V33gRNA	*GGC*UAGAACCUCUUUUAAAUUUCU ACUGUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1179	1093	405	851
IONCAS-V034	37	V34gRNA	*UUU*AAUACCUUAUAUAAAUUUCUA CUAUUGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1180	1094	406	852
IONCAS-V035	38	V35gRNA	*GUG*GAUUUGACAAACGAAAUGCCA AACUUUGAGGAA**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1195	1095	407	853
IONCAS-V036	39	V36gRNA	*GAC*UAUUAAGUCUUUUGAAUUUCU ACUGUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1189	1096	408	854
IONCAS-V037	40	V37gRNA	*GUU*UUGAAGACCAAUUAAAUUUCU ACUAUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1180	1097	409	855
IONCAS-V038	41	V38gRNA	*GUU*UAAUACCUUAUAUAAAUUUCU ACUAUUGUAGA**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1196	1098	410	856
IONCAS-V039	42	V39gRNA	*GUC*UGAAAGACUCGAAAAAUUUCU ACUGUUGUAGAU**CUGAUGGUCCA UGUCUGUUACUC***uuu*	1179	1099	411	857
IONCAS-V040	43	V40gRNA	*UUU*AGAAGAUUCUUCAAAUUUCUA CUGUUGUAGAU**CUGAUGGUCCAU GUCUGUUACUC***uuu*	1179	1100	412	858

IONCAS-V041	44	V41gRNA	*GUC*UAUAUGACUAAGUAAUUUCUACUAUGUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1197	1101	413	859
IONCAS-V042	45	V42gRNA	*GGC*AAUAAGCCAAAUGUAAUUUCUACUUUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1191	1102	414	860
IONCAS-V043	46	V43gRNA	*GUU*UAAUACCUUAUAUAAAUUUCUAUUAUUAUAGA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1198	1103	415	861
IONCAS-V044	47	V44gRNA	*GCU*ACUAAGCCUUUAUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1104	416	862
IONCAS-V045	48	V45gRNA	*GAA*GAGGUAAAGAGGCAUUCGGAAGAGGUAAAGA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1199	1105	417	863
IONCAS-V046	49	V46gRNA	*GGC*UCUAAGAGGCUCAAAUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1106	418	864
IONCAS-V047	50	V47gRNA	*GGU*UAAUAAACCCUUCUAAUUUCUACUGUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1185	1107	419	865
IONCAS-V048	51	V48gRNA	*UUU*GAGACCACCUCAUAAUUUCUACUAUCGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1200	1108	420	866
IONCAS-V049	52	V49gRNA	*GGC*UAUAAAGCCAAAAUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1109	421	867
IONCAS-V050	53	V50gRNA	*GUU*UAGCUACUCCUUAUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1110	422	868
IONCAS-V051	54	V51gRNA	*GAA*UAAUGAUUCCCUUUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1111	423	869
IONCAS-V052	55	V52gRNA	*GAA*UAAUAACCCCUUCUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1112	424	870
IONCAS-V053	56	V53gRNA	*GUC*UAUAAGGCCUUUCUAAUUUCUACUAUUGUAGAUA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1201	1113	425	871
IONCAS-V054	57	V54gRNA	*AAU*CUACAAUAGUAGAAAUUCUAUAGGCUUAUUAGCCC**UGAUGGUCCAUGUCUGUUACUC***uuu*	1202	1114	426	872
IONCAS-V055	58	V55gRNA	*GGC*UAUAAAGCCUUUAUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1115	427	873
IONCAS-V056	59	V56gRNA	*AUC*UACAAAAGUAGAAAUGGACUAUUUGUAUUUGA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1203	1116	428	874
IONCAS-V057	60	V57gRNA	*GUU*GAAUAACCUUAAAUAAUUUCUACUGUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1185	1117	429	875

IONCAS-V058	61	V58gRNA	*AGU*AACUAUAGAUAAUUUCUACUGUUGUAGAUAUAUAAA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1204	1118	430	876
IONCAS-V059	62	V59gRNA	*AUC*UACAAUAGUAGAAAUUAACUUAGGUUUAUAGCC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1205	1119	431	877
IONCAS-V060	63	V60gRNA	*GUC*UGCAGGACUAUUUAAUUUCUACUAUUAGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1206	1120	432	878
IONCAS-V061	64	V61gRNA	*CUC*UAAAGCACUGAGUAAAUUUCUACUGUUGUAGA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1207	1121	433	879
IONCAS-V062	65	V62gRNA	*GUC*UAGAGUAACAUUUAAUUUCUACUAUGUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1197	1122	434	880
IONCAS-V063	66	V63gRNA	*AUC*UACAACAGUAGAAAUUUAAUAGGUCGCCUAGAC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1208	1123	435	881
IONCAS-V064	67	V64gRNA	*GGC*UAUAAAGCCUAAAUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1124	436	882
IONCAS-V065	68	V65gRNA	*CCC*UAGAAGACCCUCCAAAUUUCUACUGUUGUAGA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1209	1125	437	883
IONCAS-V066	69	V66gRNA	*GUC*AAAGGGCACGUCCUAAUUUCUACUGUUGUAGAUA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1210	1126	438	884
IONCAS-V067	70	V67gRNA	*AUC*UACAAUAGUAGAAAUUAUUUGUGGCUUAUAGCC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1211	1127	439	885
IONCAS-V068	71	V68gRNA	*GCC*UAAUAACUAUUCAGAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1193	1128	440	886
IONCAS-V069	72	V69gRNA	*UUU*GACGAUGCCCCCCUGCGAAGCGUAAAAAUCGGUC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1212	1129	441	887
IONCAS-V070	73	V70gRNA	*AUC*UACAACAGUAGAAAUUCUCUAUAACGAUUUGGC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1213	1130	442	888
IONCAS-V071	74	V71gRNA	*CCU*AAAGCCGGCAUAAAAUUUCUACUGUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1179	1131	443	889
IONCAS-V072	75	V72gRNA	*UCU*ACAACAGUAGAAAUUUUGUAUUGGUCUUAAAC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1214	1132	444	890
IONCAS-V073	76	V73gRNA	*GUC*UAAUCCUUAUAUAAAAUUUCUACUUAUGUAGAC**UGAUGGUCCAUGUCUGUUACUC***uuu*	1215	1133	445	891
IONCAS-V074	77	V74gRNA	*GUC*UAUAGCCAAUUUAAAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1180	1134	446	892

IONCAS-V075	78	V75gRNA	*AUC*UACAAUAGUAGAAAUUGUUUUAAACGAUUAAAC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1216	1135	447	893
IONCAS-V076	79	V76gRNA	*CUC*UAAUACAUAUAGAAGAUUUCUACUUUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1217	1136	448	894
IONCAS-V077	80	V77gRNA	*GGC*UAAAGCUCUUUAAGAAUUUCUACUUUCGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1176	1061	449	895
IONCAS-V078	81	V78gRNA	*GUC*UAAACCAAACAUGAAAUUUCUACUGUCGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1218	1137	450	896
IONCAS-V079	82	V79gRNA	*GUC*CAGAGCUCAUGAUAAAUUUCUACUCUCGUAGA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1219	1138	451	897
IONCAS-V081	83	V81gRNA	*GCC*AAAUACACCUAUAAGAUUUCUACUGUUGUAG**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1220	1139	452	898
IONCAS-V082	84	V82gRNA	*UUC*AGAACUAAGAUAAAAUUUCUACUGUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1179	1140	453	899
IONCAS-V083	85	V83gRNA	*CCA*AAUAGACCCAUCAAAUUUCUACUUUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1221	1141	454	900
IONCAS-V084	86	V84gRNA	*GGC*UAAUAAGCUCUUUAAAUUUCUACUUUCGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1222	1142	455	901
IONCAS-V085	87	V85gRNA	*UCU*ACAAAAGUAGAAAUCUAUAUACAAGUAUAGACC**UGAUGGUCCAUGUCUGUUACUC***uuu*	1223	1143	456	902
IONCAS-V086	88	V86gRNA	*AGG*UACCGAAUAAAAUUUCUACUGUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1179	1144	457	903
IONCAS-V087	89	V87gRNA	*AUC*UACAACAGUAGAAAUUUAAUAUGAAGUUCAAAC**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1224	1145	458	904
IONCAS-V088	90	V88gRNA	*GUU*UCAAACCACAUUAAAAUUUCUACUGUAUGUAG**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1225	1146	459	905
IONCAS-V089	91	V89gRNA	*CCU*AAUCCAUAUAAAGAAUUUCUACUUUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1226	1147	460	906
IONCAS-V090	92	V90gRNA	*AUC*UACAAAAGUAGAAAUCUUAUAUAGGUAUUAGA**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1227	1148	461	907
IONCAS-V091	93	V91gRNA	*GGC*UAGUAAGCUCUGAUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1149	462	908
IONCAS-V092	94	V92gRNA	*GUU*UAAGAGACCCUUUUAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1178	1150	463	909

IONCAS-V093	95	V93gRNA	*GCU*GUUAAACCGUUAAAAUUUCUACUAUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1180	1151	464	910
AsCas12a	1	AsCas12agRNA	*GUC*AAAAGACCUUUUUAAUUUCUACUCUUGUAGAU**CUGAUGGUCCAUGUCUGUUACUC***uuu*	1228	1152	465	911

Guides described in the table above were used to test activity of their respective Cas proteins in a cleavage assay. The artificial synthesized DNA target is a 166 bp double stranded DNA containing a seven nucleotide PAM library followed by the target sequence. The DNA sequence (from 5' to 3') of the non-complementary strand of the synthesized DNA target is: ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTCAGCTGTTAACATCAGTACGTTA*NNNNNN***CTGATGGTCCATGTCTGTTACTC**GCCTGTCAAGTGGCGTGACACCGGGCGTGTTCCCCAGAG TGACTTAGATCGGAAGAGCACACGTCTGAACTCCAGTC (SEQ ID NO: 467), wherein the underlined, bold nucleotides represent the sequence which is the reverse complement to the target sequence, and italicized 7*N* (N = A, C, G, T) represents the PAM library.

HEK293T cells, maintained in DMEM (10% FBS, 10 U/mL penicillin, 100 µg/mL streptomycin), were seeded in 10 cm dishes (Corning) at a density of 1.5×10^6 cells/plate, and incubated for 24 hours. Cells were transfected with 5 µg/plate of pTwistCMV-puro vectors expressing Cas proteins, listed in the table above, using 20 µL BioT Transfection Reagent (Morganville Scientific). HEK293T nuclear extract was prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific) according to manufacturer's protocol. For each cleavage reaction, 5 µL of HEK293T nuclear extract expressing Cas protein and 200 nM guide RNA were incubated at room temperature with 3 µL of cleavage buffer NEBuffer r2.1 (New England BioLabs) for 10 min to facilitate guide ribonucleoprotein (RNP) complex formation. Afterwards, 10 nM of DNA template was added to the reaction for overnight incubation at 37 °C, followed by the addition of proteinase K and RNAse to clean up the reaction. DNA templates were purified using DNA Clean & Concentrator-5 (Zymo Research). Digestion of DNA templates was analyzed and visualized using an Agilent TapeStation.

Cleavage activity *in vitro* is shown in **Figure 1.**

Because Cas12a-like enzymes can sometimes share common guides, the cleavage activity screen was repeated on every Cas protein using the guide specific to AsCas12a (AsCas12agRNA) in place of their respective protein-specific guides. The results are shown in **Figure 2**.

Example 3: Identification of PAM sequences for Cas proteins

The PAM sequences of Cas proteins that displayed *in vitro* enzymatic activity were determined through dropout assay (Ran, FA et al. *Nature*. **2015**, *520*, 186-191). In the dropout assay, the uncut fraction of

the DNA templates was amplified via PCR reaction using i5/i7 index primers from Integrated DNA Technologies (IDT). Amplicons were sequenced via Illumina MiSeq. For a given single-end MiSeq sequencing run, data was demultiplexed to generate FASTQ files using either MiSeq machine or Illumina bcl2fastq Conversion Software (support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software/downloads.html). A custom program was written to extract 7-mer PAM sequences located 5' to a known spacer sequence from the FASTQ files. For each extracted PAM sequence in a sample, a count-per-million (CPM) value was calculated by dividing the number of reads containing the PAM sequence by the total number of reads of the sample. Relative PAM sequence depletion frequencies were calculated by dividing CPM values in a Cas protein treated sample by the corresponding CPM values in an untreated control sample. PAM sequence motifs at different depletion frequency thresholds were then generated using Python Logomaker software library available at logomaker.readthedocs.io/en/latest/ (Tareen and Kinney 2019). The consensus PAM sequences of Cas proteins described above are summarized in the table below.

Table 6

Consensus PAM sequences of Cas proteins

Cas Protein ID	Cas Protein SEQ ID NO.	PAM Sequence
IONCAS-V021	24	YTTV
IONCAS-V023	26	YTTV
IONCAS-V024	27	YYYM
IONCAS-V025	28	YYTC
IONCAS-V026	29	TTTM
IONCAS-V027	30	RTTT
IONCAS-V029	32	TTTV
IONCAS-V031	34	TTTR
IONCAS-V034	37	YTTV
IONCAS-V035	38	TTTV
IONCAS-V039	42	TTTV
IONCAS-V042	45	TTTV
IONCAS-V043	46	YYTV
IONCAS-V044	47	TTTV
IONCAS-V047	50	YTS
IONCAS-V050	53	TTTS
IONCAS-V051	54	TTTV
IONCAS-V055	58	TTTV

IONCAS-V071	74	YYYV
IONCAS-V073	76	YTV
IONCAS-V082	85	TTTV
IONCAS-V085	88	YYYV
IONCAS-V088	91	TTTV
AsCas12a	1	TTTV

In the table above, "Y" represents C or T; "M" represents A or C; "R" represents A or G; "S" represents G or C; and "V" represents A, C, or G.

Example 4: Evaluation of Cas protein activity in HEK293 cells

The enzymatic activities of Cas proteins described above were evaluated in HEK293 cells, measuring editing in nine genomic regions. For each Cas protein and each genomic target, a guide was designed comprising a target-recognition region sequence linked to the 3'-end of a protein-recognition region sequence.

Nine genomic regions were selected as targets for evaluation of cellular Cas protein activity. These regions are located within human genomic loci of DNMT1 (site 1, site 3), NLRC4, RUNX1 (site 2, site 3, site 4), AGBL1 (matched site 1), and FANCF (site 1, site 2). For each Cas protein and target, target-recognition regions that are 23 nucleotides in length were designed using CHOPCHOP v3 (Labun *et al. Nucleic Acids Res.* 2019, *47*(W1): W171-W174) and Benchling software, or sourced from existing publications (Luk *et al.*, *GEN Biotechnol.* 2022, *1*(3): 271-284; Moon *et al.*, *Nat Commun.* 2018, *9*(1):3651; Kleinstiver, B. *et al.*, *Nat Biotechnol.* 2016, *34*(8):869–874). Guides were then designed by appending the Cas protein-specific target-recognition region sequence to the 3′-end of a protein-recognition sequence described herein above (SEQ ID NO: 1061-1152). Based on the screening results shown in **Figures 2A** and **2B**, guides for Cas proteins IONCAS-V023, -V031, -V035, -V043, -V047, -V073, -V082, -V085, -V086, and -V088 were designed using the same protein recognition region as AsCas12a, and therefore have the same guide sequence. Guide sequences for each Cas protein and target are presented in the sequence listing, as indicated by the corresponding SEQ ID NO in the tables below.

2×10^5 HEK293 cells were transfected with 400 ng/well of codon-optimized DNA encoding the Cas protein and 400 ng/well of guide expressing plasmid using 4D Nucleofector (Lonza) and cells were seeded into 96-well plates after transfection. After a 72-hour incubation period, genomic DNA was extracted from cells using DNeasy Blood & Tissue kits (Qiagen) following the manufacturer's recommended protocol. Genomic PCRs were performed using primers spanning the target region, and libraries were constructed using i5/i7 index primers from IDT. Amplicons were sequenced via MiSeq (Illumina). For a given paired-end MiSeq sequencing run, data was demultiplexed to generate FASTQ files as described above. The

CRISPRessoBatch module of CRISPResso2 software (Clement et al. *Nat Biotechnol,* 2019) was used to process and analyze all sample sequencing data. A custom script was written to calculate editing percentages based on InDels only from CRISPResso2 analysis results. The gene editing quantification window was set to be ~40 nucleotides long centered on a guide RNA sequence of ~20 nucleotides, with ~10 nucleotides upstream and ~10 nucleotides downstream. To enable visualization alignments of low frequency InDels reads, the minimum threshold for report reads was set to 0.01% (default is 0.20%) and maximum reported alignments was set to 100 (default is 50). Results are presented in the table below as percentage of reads with InDels (%InDels) relative to the total number of reads within the 40 nucleotide window.

In the tables below, results are grouped by genomic target, as indicated in the table title. For each Cas editing system that was tested, the Cas protein ID and its SEQ ID NO. are indicated in the columns under "Cas Protein", and the Guide ID and its SEQ ID NO. are indicated in the columns under "Guide."

Table 7

% InDels of DNMT1 Site 1 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	**Cas Protein SEQ ID NO**	**Guide ID**	**Guide SEQ ID NO**	
IONCAS-V021	24	gV21-DNM-1	611	1.76
IONCAS-V024	27	gV24-DNM-1	612	2.69
IONCAS-V025	28	gV25-DNM-1	613	0.10
IONCAS-V026	29	gV26-DNM-1	614	0.17
IONCAS-V027	30	gV27-DNM-1	615	0.05
IONCAS-V029	32	gV29-DNM-1	616	0.63
IONCAS-V034	37	gV34-DNM-1	617	0.07
IONCAS-V039	42	gV39-DNM-1	618	0.55
IONCAS-V042	45	gV42-DNM-1	619	0.13
IONCAS-V044	47	gV44-DNM-1	620	0.66
IONCAS-V050	53	gV50-DNM-1	621	0.97
IONCAS-V051	54	gV51-DNM-1	622	1.67
IONCAS-V055	58	gV55-DNM-1	623	1.05
IONCAS-V071	74	gV71-DNM-1	624	4.51
IONCAS-V023	26	V23gAS-DNM-1	625	0.56
IONCAS-V031	34	V31gAS-DNM-1	625	0.15
IONCAS-V035	38	V35gAS-DNM-1	625	0.81
IONCAS-V043	46	V43gAS-DNM-1	625	0.76
IONCAS-V047	50	V47gAS-DNM-1	625	0.14
IONCAS-V073	76	V73gAS-DNM-1	625	0.16
IONCAS-V082	84	V82gAS-DNM-1	625	0.73
IONCAS-V085	87	V85gAS-DNM-1	625	0.14

IONCAS-V088	90	V88gAS-DNM-1	625	0.48
IONCAS-V086	88	V86gAS-DNM-1	625	0.53
AsCas12a	1	gAS-DNM-1	625	3.61

Table 8

% InDels of DNMT1 Site 3 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	Cas Protein SEQ ID NO	Guide ID	Guide SEQ ID NO	
IONCAS-V021	24	gV21-DNM-3	626	10.19
IONCAS-V024	27	gV24-DNM-3	627	63.44
IONCAS-V025	28	gV25-DNM-3	628	1.21
IONCAS-V026	29	gV26-DNM-3	629	1.78
IONCAS-V027	30	gV27-DNM-3	630	0.21
IONCAS-V029	32	gV29-DNM-3	631	6.94
IONCAS-V034	37	gV34-DNM-3	632	0.12
IONCAS-V039	42	gV39-DNM-3	633	4.22
IONCAS-V042	45	gV42-DNM-3	634	0.23
IONCAS-V044	47	gV44-DNM-3	635	8.01
IONCAS-V050	53	gV50-DNM-3	636	20.98
IONCAS-V051	54	gV51-DNM-3	637	33.77
IONCAS-V055	58	gV55-DNM-3	638	5.54
IONCAS-V071	74	gV71-DNM-3	639	62.74
IONCAS-V023	26	V23gAS-DNM-3	640	2.53
IONCAS-V031	34	V31gAS-DNM-3	640	0.23
IONCAS-V035	38	V35gAS-DNM-3	640	14.69
IONCAS-V043	46	V43gAS-DNM-3	640	17.77
IONCAS-V047	50	V47gAS-DNM-3	640	1.97
IONCAS-V073	76	V73gAS-DNM-3	640	0.59
IONCAS-V082	84	V82gAS-DNM-3	640	22.07
IONCAS-V085	87	V85gAS-DNM-3	640	0.65
IONCAS-V088	90	V88gAS-DNM-3	640	4.99
IONCAS-V086	88	V86gAS-DNM-3	640	3.76
AsCas12a	1	gAS-DNM-3	640	33.27

Table 9

% InDels of FANCF Site 1 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	Cas Protein SEQ ID NO	Guide ID	Guide SEQ ID NO	
IONCAS-V021	24	gV21-FA-1	641	31.87
IONCAS-V025	28	gV25-FA-1	642	0.14
IONCAS-V027	30	gV27-FA-1	643	0.04
IONCAS-V029	32	gV29-FA-1	644	33.12
IONCAS-V034	37	gV34-FA-1	645	4.50
IONCAS-V039	42	gV39-FA-1	646	11.47
IONCAS-V042	45	gV42-FA-1	647	1.74
IONCAS-V044	47	gV44-FA-1	648	36.84
IONCAS-V050	53	gV50-FA-1	649	47.93
IONCAS-V051	54	gV51-FA-1	650	35.87
IONCAS-V055	58	gV55-FA-1	651	42.96
IONCAS-V071	74	gV71-FA-1	652	54.48
IONCAS-V023	26	V23gAS-FA-1	653	15.74
IONCAS-V031	34	V31gAS-FA-1	653	0.89
IONCAS-V035	38	V35gAS-FA-1	653	30.76
IONCAS-V043	46	V43gAS-FA-1	653	37.68
IONCAS-V047	50	V47gAS-FA-1	653	7.33
IONCAS-V073	76	V73gAS-FA-1	653	0.91
IONCAS-V082	84	V82gAS-FA-1	653	39.03
IONCAS-V085	87	V85gAS-FA-1	653	1.12
IONCAS-V088	90	V88gAS-FA-1	653	49.01
IONCAS-V086	88	V86gAS-FA-1	653	12.36
AsCas12a	1	gAS-FA-1	653	39.71

Table 10

% InDels of FANCF Site 2 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	Cas Protein SEQ ID NO	Guide ID	Guide SEQ ID NO	
IONCAS-V021	24	gV21-FA-2	654	4.37
IONCAS-V025	28	gV25-FA-2	655	0
IONCAS-V027	30	gV27-FA-2	656	0
IONCAS-V029	32	gV29-FA-2	657	0
IONCAS-V034	37	gV34-FA-2	658	1
IONCAS-V039	42	gV39-FA-2	659	0

IONCAS-V042	45	gV42-FA-2	660	10.00
IONCAS-V044	47	gV44-FA-2	661	11.11
IONCAS-V050	53	gV50-FA-2	662	1.22
IONCAS-V051	54	gV51-FA-2	663	4.10
IONCAS-V055	58	gV55-FA-2	664	0
IONCAS-V071	74	gV71-FA-2	665	0
IONCAS-V023	26	V23gAS-FA-2	666	15.05
IONCAS-V031	34	V31gAS-FA-2	666	0.59
IONCAS-V035	38	V35gAS-FA-2	666	18.23
IONCAS-V043	46	V43gAS-FA-2	666	8.45
IONCAS-V047	50	V47gAS-FA-2	666	0
IONCAS-V073	76	V73gAS-FA-2	666	0.39
IONCAS-V082	84	V82gAS-FA-2	666	0
IONCAS-V085	87	V85gAS-FA-2	666	31.79
IONCAS-V088	90	V88gAS-FA-2	666	0.89
IONCAS-V086	88	V86gAS-FA-2	666	10.49
AsCas12a	1	gAS-FA-2	666	26.42

Table 11

% InDels of NLRC4 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	Cas Protein SEQ ID NO	Guide ID	Guide SEQ ID NO	
IONCAS-V021	24	gV21-NLR	667	35.7
IONCAS-V024	27	gV24-NLR	668	54.22
IONCAS-V025	28	gV25-NLR	669	0.63
IONCAS-V026	29	gV26-NLR	670	3.23
IONCAS-V027	30	gV27-NLR	671	0.09
IONCAS-V029	32	gV29-NLR	672	30.21
IONCAS-V034	37	gV34-NLR	673	2.27
IONCAS-V039	42	gV39-NLR	674	25.07
IONCAS-V042	45	gV42-NLR	675	0.22
IONCAS-V044	47	gV44-NLR	676	48
IONCAS-V051	54	gV51-NLR	677	48.34
IONCAS-V055	58	gV55-NLR	678	43.56
IONCAS-V071	74	gV71-NLR	679	61.29
IONCAS-V023	26	V23gAS-NLR	680	23.61
IONCAS-V031	34	V31gAS-NLR	680	0.5
IONCAS-V035	38	V35gAS-NLR	680	43.67
IONCAS-V043	46	V43gAS-NLR	680	9.89
IONCAS-V047	50	V47gAS-NLR	680	1.55
IONCAS-V073	76	V73gAS-NLR	680	0.72

IONCAS-V082	84	V82gAS-NLR	680	35.17
IONCAS-V085	87	V85gAS-NLR	680	1.08
IONCAS-V088	90	V88gAS-NLR	680	38.66
AsCas12a	1	gAS-NLR	680	32.17

Table 12

% InDels of AGBL1 matched site 1 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	Cas Protein SEQ ID NO	Guide ID	Guide SEQ ID NO	
IONCAS-V021	24	gV21-MS-1	681	0.48
IONCAS-V024	27	gV24-MS-1	682	0.12
IONCAS-V025	28	gV25-MS-1	683	0.12
IONCAS-V026	29	gV26-MS-1	684	0.12
IONCAS-V027	30	gV27-MS-1	685	0.2
IONCAS-V029	32	gV29-MS-1	686	0.17
IONCAS-V034	37	gV34-MS-1	687	19.45
IONCAS-V039	42	gV39-MS-1	688	2.72
IONCAS-V042	45	gV42-MS-1	689	8.44
IONCAS-V044	47	gV44-MS-1	690	2.78
IONCAS-V055	58	gV55-MS-1	691	0.18
IONCAS-V071	74	gV71-MS-1	692	0.97
IONCAS-V023	26	V23gAS-MS-1	693	19.29
IONCAS-V031	34	V31gAS-MS-1	693	0.17
IONCAS-V035	38	V35gAS-MS-1	693	0.44
IONCAS-V043	46	V43gAS-MS-1	693	0.16
IONCAS-V047	50	V47gAS-MS-1	693	0
IONCAS-V073	76	V73gAS-MS-1	693	0.11
IONCAS-V082	84	V82gAS-MS-1	693	41.39
IONCAS-V085	87	V85gAS-MS-1	693	8.59
IONCAS-V088	90	V88gAS-MS-1	693	0.63
IONCAS-V086	88	V86gAS-MS-1	693	2.78
AsCas12a	1	gAS-MS-1	693	3.13

Table 13

% InDels of RUNX1 Site 2 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	Cas Protein SEQ ID NO	Guide ID	Guide SEQ ID NO	
IONCAS-V021	24	gV21-RUN-2	694	2.99

IONCAS-V024	27	gV24-RUN-2	695	18.74
IONCAS-V025	28	gV25-RUN-2	696	0.10
IONCAS-V026	29	gV26-RUN-2	697	0.55
IONCAS-V027	30	gV27-RUN-2	698	0.13
IONCAS-V029	32	gV29-RUN-2	699	0.23
IONCAS-V034	37	gV34-RUN-2	700	0.29
IONCAS-V039	42	gV39-RUN-2	701	0.76
IONCAS-V042	45	gV42-RUN-2	702	0.32
IONCAS-V044	47	gV44-RUN-2	703	0.82
IONCAS-V050	53	gV50-RUN-2	704	0.43
IONCAS-V051	54	gV51-RUN-2	705	24.5
IONCAS-V055	58	gV55-RUN-2	706	0.44
IONCAS-V071	74	gV71-RUN-2	707	15.57
IONCAS-V023	26	V23gAS-RUN-2	708	0.45
IONCAS-V031	34	V31gAS-RUN-2	708	0.36
IONCAS-V035	38	V35gAS-RUN-2	708	1.24
IONCAS-V043	46	V43gAS-RUN-2	708	3.70
IONCAS-V047	50	V47gAS-RUN-2	708	0.59
IONCAS-V073	76	V73gAS-RUN-2	708	1.42
IONCAS-V082	84	V82gAS-RUN-2	708	7.91
IONCAS-V085	87	V85gAS-RUN-2	708	0.93
IONCAS-V088	90	V88gAS-RUN-2	708	9.78
IONCAS-V086	88	V86gAS-RUN-2	708	7.04
AsCas12a	1	gAS-RUN-2	708	43.54

Table 14

% InDels of RUNX1 Site 3 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	**Cas Protein SEQ ID NO**	**Guide ID**	**Guide SEQ ID NO**	
IONCAS-V021	24	gV21-RUN-3	709	13.65
IONCAS-V024	27	gV24-RUN-3	710	66.76
IONCAS-V025	28	gV25-RUN-3	711	0.62
IONCAS-V026	29	gV26-RUN-3	712	6.59
IONCAS-V027	30	gV27-RUN-3-4	713	0.36
IONCAS-V029	32	gV29-RUN-3	714	12.31
IONCAS-V034	37	gV34-RUN-3	715	0.78
IONCAS-V039	42	gV39-RUN-3	716	6.66
IONCAS-V042	45	gV42-RUN-3	717	0.62
IONCAS-V044	47	gV44-RUN-3	718	20.73
IONCAS-V050	53	gV50-RUN-3	719	21.89
IONCAS-V051	54	gV51-RUN-3	720	19.86

IONCAS-V055	58	gV55-RUN-3	721	9.35
IONCAS-V071	74	gV71-RUN-3	722	59.07
IONCAS-V023	26	V23gAS-RUN-3	723	3.77
IONCAS-V031	34	V31gAS-RUN-3	723	0.17
IONCAS-V035	38	V35gAS-RUN-3	723	13.35
IONCAS-V043	46	V43gAS-RUN-3	723	36.47
IONCAS-V047	50	V47gAS-RUN-3	723	2.14
IONCAS-V073	76	V73gAS-RUN-3	723	0.42
IONCAS-V082	84	V82gAS-RUN-3	723	8.02
IONCAS-V085	87	V85gAS-RUN-3	723	1.17
IONCAS-V088	90	V88gAS-RUN-3	723	12.82
IONCAS-V086	88	V86gAS-RUN-3	723	7.41
AsCas12a	1	gAS-RUN-3	723	58.52

Table 15

% InDels of RUNX1 Site 4 using Cas proteins and guides in HEK293 cells

Cas Protein		Guide		%InDels
Cas Protein ID	**Cas Protein SEQ ID NO**	**Guide ID**	**Guide SEQ ID NO**	
IONCAS-V021	24	gV21-RUN-4	724	21.2
IONCAS-V024	27	gV24-RUN-4	725	53.81
IONCAS-V025	28	gV25-RUN-4	726	0.34
IONCAS-V026	29	gV26-RUN-4	727	1.31
IONCAS-V029	32	gV29-RUN-4	728	7.43
IONCAS-V034	37	gV34-RUN-4	729	0.71
IONCAS-V039	42	gV39-RUN-4	730	6.00
IONCAS-V042	45	gV42-RUN-4	731	0.51
IONCAS-V044	47	gV44-RUN-4	732	4.65
IONCAS-V050	53	gV50-RUN-4	733	17.17
IONCAS-V051	54	gV51-RUN-4	734	8.56
IONCAS-V055	58	gV55-RUN-4	735	7.73
IONCAS-V071	74	gV71-RUN-4	736	63.12
IONCAS-V023	26	V23gAS-RUN-4	737	1.58
IONCAS-V031	34	V31gAS-RUN-4	737	0.34
IONCAS-V035	38	V35gAS-RUN-4	737	1.61
IONCAS-V043	46	V43gAS-RUN-4	737	11.58
IONCAS-V047	50	V47gAS-RUN-4	737	1.28
IONCAS-V073	76	V73gAS-RUN-4	737	0.50
IONCAS-V082	84	V82gAS-RUN-4	737	18.16
IONCAS-V085	87	V85gAS-RUN-4	737	0.86
IONCAS-V088	90	V88gAS-RUN-4	737	16.38
IONCAS-V086	88	V86gAS-RUN-4	737	4.72

AsCas12a	1	gAS-RUN-4	737	50.27

Example 5: Effect of target-recognition region length on Cas protein activity in HepG2 cells, LNP delivery

Modified guides comprising a target-recognition region linked to the 3'-end of a protein-recognition region were designed for selected Cas proteins from the example above. The guides were optimized by tuning the target-recognition region length, and then screened for their effect on the enzymatic activity of their respective Cas proteins in an *in vitro* assay wherein HepG2 cells were incubated with an LNP formulation of a Cas protein and guide editing system.

Guides were designed and synthesized comprising a protein-recognition region having a 20 nucleotide direct repeat region appended to two 5'-terminal nucleotides, CU; a target-recognition region with varying lengths of 20-24 nucleotides; and three 3'-terminal nucleotides, UUU. Guides designed for IONCAS-V035 and IONCAS-V082 share a protein-recognition sequence, and guides designed for IONCAS-V050 and IONCAS-V051 share a protein recognition sequence. The sequence of each guide is provided in the sequence listing according to the SEQ ID NO. indicated in the tables below. Each guide in the tables below has the following modifications: the nucleosides at positions 1 and 2 at the 5'-end are 2'-OMe sugar moieties wherein each nucleoside is linked to the next nucleoside with a phosphorothioate internucleoside linkage; and the last two nucleosides at the 3'-end are 2'-OMe sugar moieties, wherein each nucleoside is linked to the preceding nucleoside by a phosphorothioate internucleoside linkage.

Human HepG2 cells were split at the density of 19,000 cells per well in 96-well plates and cultured in DMEM medium with 10% FBS and 1% PenStrep. RNA/LNP complexes were generated by mixing RNA (Cas protein mRNA / guide at 1:2 ratio by weight) with a LNP formulation (Kazemian, et al., *Mol. Pharmaceutics*, 2022, 19, 1669-1689; Albertsen et al., *Adv. Drug Deliv. Rev.* 2022, 118, 114416; Cullis et al., *Nat. Rev. Drug Discov.* 2024, 23, 709-722.). After incubating for 15 minutes, the mixture was added to the culture medium at total Cas protein mRNA + guide RNA doses indicated in the tables below. Cells were harvested for InDel analysis after 72 hours, and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below. "N.D." indicates that the data point was not determined.

In the tables below, results are grouped by Cas protein as indicated in the table title and in the "Cas Protein ID" column. Results are further grouped by target, as indicated in the "Target" column. For each Cas protein and guide editing system that was tested, the unmodified guide sequences and chemically-modified guide sequences are provided in the sequence listing as indicated in the table below under the columns "Guide SEQ ID NO.", and "Mod. Guide SEQ ID NO." respectively.

Table 16

Effect of guide target-recognition region length on % InDels of target genes using IONCAS-V035

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Target Recognition Region Length	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V035	38	1859672	20	738	912	44.70	1.92	FANCF Site 1
IONCAS-V035	38	1859673	21	739	913	45.97	0.50	
IONCAS-V035	38	1859674	22	740	914	35.02	0.39	
IONCAS-V035	38	1859675	23	741	915	40.23	0.32	
IONCAS-V035	38	1859676	24	742	916	35.12	1.63	
IONCAS-V035	38	1859687	20	743	917	90.97	9.43	NLRC4
IONCAS-V035	38	1859688	21	744	918	65.04	1.14	
IONCAS-V035	38	1859689	22	745	919	91.01	3.01	
IONCAS-V035	38	1859690	23	746	920	91.31	10.23	
IONCAS-V035	38	1859691	24	747	921	87.42	4.64	

Table 17

Effect of guide target-recognition region length on % InDels of target genes using IONCAS-V050

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Target Recognition Region Length	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V050	53	1859662	20	748	922	50.72	0.74	FANCF Site 1
IONCAS-V050	53	1859663	21	749	923	35.60	1.28	
IONCAS-V050	53	1859664	22	750	924	39.21	0.74	
IONCAS-V050	53	1859665	23	751	925	34.95	0.49	
IONCAS-V050	53	1859666	24	752	926	25.10	0.35	
IONCAS-V050	53	1859682	20	753	927	1.14	0.08	RUNX1 Site 3
IONCAS-V050	53	1859683	21	754	928	0.83	0.17	
IONCAS-V050	53	1859684	22	755	929	0.46	N.D.	
IONCAS-V050	53	1859685	23	756	930	0.42	0.08	
IONCAS-V050	53	1859686	24	757	931	0.50	0.06	

Table 18

Effect of guide target-recognition region length on % InDels of target genes using IONCAS-V051

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Target Recognition Region Length	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V051	54	1859662	20	748	922	19.66	0.43	FANCF Site 1
IONCAS-V051	54	1859663	21	749	923	12.58	0.21	
IONCAS-V051	54	1859664	22	750	924	14.19	0.25	
IONCAS-V051	54	1859665	23	751	925	14.76	0.91	
IONCAS-V051	54	1859666	24	752	926	14.31	0.46	
IONCAS-V051	54	1859667	20	758	932	67.18	1.10	NLRC4
IONCAS-V051	54	1859668	21	759	933	51.77	1.56	
IONCAS-V051	54	1859669	22	760	934	71.28	2.41	
IONCAS-V051	54	1859670	23	761	935	91.37	3.39	
IONCAS-V051	54	1859671	24	762	936	92.37	9.73	

Table 19

Effect of guide target-recognition region length on % InDels of target genes using IONCAS-V071

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Target Recognition Region Length	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V071	74	1859339	20	763	937	5.57	0.28	DNMT1 Site 3
IONCAS-V071	74	1859340	21	764	938	9.89	0.39	
IONCAS-V071	74	1859341	22	765	939	14.33	0.31	
IONCAS-V071	74	1859342	23	766	940	12.15	0.27	
IONCAS-V071	74	1859343	24	767	941	17.34	0.67	
IONCAS-V071	74	1859344	20	768	942	11.16	0.66	RUNX1 Site 4
IONCAS-V071	74	1859345	21	769	943	3.84	0.36	
IONCAS-V071	74	1859346	22	770	944	4.56	0.28	
IONCAS-V071	74	1859660	23	771	945	2.19	0.07	
IONCAS-V071	74	1859661	24	772	946	2.05	0.23	

Table 20

Effect of guide target-recognition region length on % InDels of target genes using IONCAS-V082

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Target Recognition Region Length	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V082	84	1859672	20	738	912	73.26	1.57	FANCF Site 1
IONCAS-V082	84	1859673	21	739	913	61.31	0.64	
IONCAS-V082	84	1859674	22	740	914	43.65	0.56	
IONCAS-V082	84	1859675	23	741	915	48.78	1.38	
IONCAS-V082	84	1859676	24	742	916	54.00	0.47	
IONCAS-V082	84	1859677	20	773	947	1.80	0.06	AGBL1 matched site 1
IONCAS-V082	84	1859678	21	774	948	1.32	0.02	
IONCAS-V082	84	1859679	22	775	949	1.19	0.08	
IONCAS-V082	84	1859680	23	776	950	0.47	0.10	
IONCAS-V082	84	1859681	24	777	951	0.24	0.05	

Guides were designed with a protein recognition region comprising a 20-nucleotide direct repeat and a 5'-terminal UU (IONCAS-V071) or UA (IONCAS-V051); a target-recognition region with a length of 20 to 24 nucleotides; and a three nucleotide 3'-terminal UUU. The sequence of each guide is provided in the sequence listing according to the SEQ ID NO. indicated in the tables below. Each guide in the tables below has the following modifications: the nucleosides at positions 1 and 2 at the 5'-end are 2'-OMe sugar moieties wherein each nucleoside is linked to the next nucleoside with a phosphorothioate internucleoside linkage; and the last two nucleosides at the 3'-end are 2'-OMe sugar moieties, wherein each nucleoside is linked to the preceding nucleoside by a phosphorothioate internucleoside linkage.

Human HepG2 cells were split at the density of 18,000 cells per well in 96-well plates and cultured in DMEM medium with 10% FBS and 1% PenStrep. After 24 hours, culture media were changed. RNA/LNP complexes were made as described herein above and were added to the culture medium at total Cas protein mRNA + guide RNA doses indicated in the tables below. Cells were harvested for InDel analysis after 72 hours, and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below.

In the tables below, results are grouped by Cas protein as indicated in the table title and in the "Cas Protein ID" column. Results are further grouped by target, as indicated in the "Target" column. For each Cas protein and guide editing system that was tested, the unmodified guide sequences and chemically-modified guide sequences are provided in the sequence listing as indicated in the table below under the columns "Guide SEQ ID NO.", and "Mod. Guide SEQ ID NO." respectively.

Table 21

Effect of guide target-recognition region length on % InDels of gene targets using IONCAS-V051

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Target Recognition Region Length	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V051	54	1886986	20	993	1027	1.12	0.13	DNMT1 Site 3
IONCAS-V051	54	1886987	21	994	1028	3.88	0.68	
IONCAS-V051	54	1886988	22	995	1029	2.68	0.30	
IONCAS-V051	54	1886989	23	996	1030	1.90	0.41	
IONCAS-V051	54	1886990	24	997	1031	1.81	0.50	
IONCAS-V051	54	1886991	20	998	1032	0.18	0.09	RUNX1 Site 2
IONCAS-V051	54	1886992	21	999	1033	0.26	0.22	
IONCAS-V051	54	1886993	22	1000	1034	0.30	0.11	
IONCAS-V051	54	1886994	23	1001	1035	0.22	0.18	
IONCAS-V051	54	1886995	24	1002	1036	0.19	0.17	

Table 22

Effect of guide target-recognition region length on % InDels of gene targets using IONCAS-V071

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Target Recognition Region Length	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V071	74	1886976	20	1003	1037	9.97	11.75	RUNX1 Site 3
IONCAS-V071	74	1886977	21	1004	1038	7.32	6.03	
IONCAS-V071	74	1886978	22	1005	1039	11.70	4.77	
IONCAS-V071	74	1886979	23	1006	1040	7.72	5.59	
IONCAS-V071	74	1886980	24	1007	1041	5.76	4.78	
IONCAS-V071	74	1886981	20	1008	1042	83.27	26.32	NLRC4
IONCAS-V071	74	1886982	21	1009	1043	89.59	31.12	
IONCAS-V071	74	1886983	22	1010	1044	92.08	39.02	
IONCAS-V071	74	1886984	23	1011	1045	94.19	33.01	
IONCAS-V071	74	1886985	24	1012	1046	96.56	36.88	

Example 6: Effect of 5'-stabilizing region on Cas protein activity in HepG2 cells, LNP delivery

Modified guides comprising a target-recognition region linked to the 3'-end of a protein-recognition region were designed for Cas proteins selected from the example above. The guides were optimized by

varying the length of the protein-recognition region by appending a 5'-stabilizing region to the direct repeat region, and then screened for their effect on the enzymatic activity of their respective Cas proteins in a HepG2 *in vitro* assay described herein above.

Guides were designed and synthesized comprising a protein-recognition region having a direct repeat region; a 23-nucleotide target-recognition region; a three nucleotide UUU 3'-extension, and optionally, a 5'-stabilizing region. The direct repeat region of each guide is 20 nucleotides. The 5'-stabilizing region is 15 or 16 nucleotides long, as reflected in the guide sequences and indicated in the table below, and consists of the same 5'-extension described herein above (SEQ ID NO: 1172-1175). The sequences of the protein-recognition regions for IONCAS-V035 and IONCAS-082 are identical.

The sequence of each guide is provided in the sequence listing according to the SEQ ID NO. indicated in the tables below. Each guide in the tables below has the following modifications: the nucleosides at positions 1 and 2 at the 5'-end are 2'-OMe sugar moieties wherein each nucleoside is linked to the next nucleoside with a phosphorothioate internucleoside linkage; and the last two nucleosides at the 3'-end are 2'-OMe sugar moieties, wherein each nucleoside is linked to the preceding nucleoside by a phosphorothioate internucleoside linkage.

Human HepG2 cells were split at the density of 19,000 cells per well in 96-well plates and cultured as described herein above. RNA/LNP complexes were generated and added as described herein above to the culture medium at total Cas protein mRNA + guide RNA doses indicated in the tables below. Cells were harvested for InDel analysis after 72 hours, and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below.

In the tables below, results are grouped by Cas protein as indicated by the table title and in the "Cas Protein ID" column. Results are further grouped by target, as indicated in the "Target" column. For each Cas protein and guide editing system that was tested, the unmodified guide sequences and chemically-modified guide sequences are provided in the sequence listing as indicated in the table below under the columns "Guide SEQ ID NO.", and "Mod. Guide SEQ ID NO." respectively.

Table 23

Effect of 5'-stabilizing region on % InDels of gene targets using IONCAS-V035

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5'-stabilizing region	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V035	38	1859332	15 nt.	812	986	57.43	4.95	FANCF Site 1
IONCAS-V035	38	1859672	none	738	912	40.23	0.32	

IONCAS-V035	38	1859336	15 nt.	813	987	95.04	15.04	NLRC4	
IONCAS-V035	38	1859687	none	743	917	91.31	10.23		

Table 24

Effect of 5'-stabilizing region on % InDels of gene targets using IONCAS-V050

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5'-stabilizing region	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V050	53	1859334	16 nt.	814	988	52.25	1.79	FANCF Site 4
IONCAS-V050	53	1859662	none	748	922	34.95	0.49	
IONCAS-V050	53	1859335	16 nt.	815	989	0.33	0.16	RUNX1 Site 3
IONCAS-V050	53	1859682	none	753	927	0.42	0.08	

Table 25

Effect of 5'-stabilizing region on % InDels of gene targets using IONCAS-V051

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5'-stabilizing region	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V051	54	1859330	16 nt.	816	990	24.04	2.09	FANCF Site 1
IONCAS-V051	54	1859662	none	748	922	14.76	0.91	
IONCAS-V051	54	1859331	16 nt.	817	991	94.48	11.45	NLRC4
IONCAS-V051	54	1859667	none	758	932	91.37	3.39	

Table 26

Effect of 5'-stabilizing region on % InDels of gene targets using IONCAS-V071

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5'-stabilizing region	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V071	74	1859328	15 nt.	778	952	34.74	2.46	DNMT1 Site 3
IONCAS-V071	74	1859339	none	763	937	12.15	0.27	
IONCAS-V071	74	1859329	15 nt.	785	959	45.64	5.85	RUNX1 Site 4
IONCAS-V071	74	1859344	none	768	942	2.19	0.07	

Table 27

Effect of 5'-stabilizing region on % InDels of gene targets using IONCAS-V082

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5'-stabilizing region	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V082	84	1859332	15 nt.	812	986	79.98	1.91	FANCF Site 1
IONCAS-V082	84	1859672	none	738	912	48.78	1.38	
IONCAS-V082	84	1859333	15 nt.	818	992	0.41	0.05	AGBL1 Matched Site 1
IONCAS-V082	84	1859677	none	773	947	0.47	0.10	

Table 28

Effect of 5'-stabilizing region on % InDels of gene targets using AsCas12a

Cas Protein		Guide				%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5'-stabilizing region	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
AsCas12a	1	1859156	15 nt.	792	966	64.97	13.13	RUNX1 Site 3
AsCas12a	1	1859337	none	793	967	28.60	1.72	
AsCas12a	1	1859327	15 nt.	800	974	92.08	10.19	RUNX1 Site 4
AsCas12a	1	1859338	none	801	975	11.78	1.21	

Example 7: Effect of guide stabilizing regions on Cas protein activity in HepG2 cells, LNP delivery

Modified guides comprising stabilizing regions appended at the 5'- or 3'-end were designed for IONCAS-V071 or *As*Cas12a. The guides were screened for their effect on the enzymatic activity of their respective Cas proteins in an *in vitro* assay wherein HepG2 cells were incubated with an LNP formulation of a Cas protein and guide editing system.

Guides were designed comprising a protein recognition region having a 20 nucleotide direct repeat region; a 23 nucleotide target-recognition region; and a three nucleotide UUU 3'-extension. Some guides were designed with a stabilizing region selected from Table 2 appended with a poly-U linker either 5' of the direct repeat (in the protein-recognition region) or 3' of the target-recognition region (in the 3'-extension region). Guides that were designed without a 5'- stabilizing region in their protein-recognition regions instead have two 5'-terminal CU nucleotides. As a control, guides with no stabilizing region and 20 nucleotide direct repeats in their protein-recognition regions were designed and tested. Guides with a 15-16 n.t. 5'-stabilizing region in their protein-recognition regions were also included.

The sequence of each guide is provided in the sequence listing according to the SEQ ID NO. indicated in the tables below. Each guide in the tables below has the following modifications: the nucleosides at positions 1 and 2 at the 5'-end are 2'-OMe sugar moieties wherein each nucleoside is linked to the next nucleoside with phosphorothioate internucleoside linkage; and the last two nucleosides at the 3'-end are 2'-OMe sugar moieties, wherein each nucleoside is linked to the preceding nucleoside by a phosphorothioate internucleoside linkage.

Human HepG2 cells were split at the density of 19,000 cells per well in 96-well plates and cultured as described herein above. RNA/LNP complexes were generated by mixing RNA with a LNP formulation, as described herein above. After incubating for 15 minutes, the mixture was added to the culture medium at total Cas protein mRNA + guide RNA doses indicated in the tables below. Cells were harvested for InDel analysis after 72 hours, and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below.

In the tables below, results are grouped by Cas protein, as indicated in the "Cas Protein ID" column, and by target, as indicated in the "Target" column. For each Cas protein and guide editing system that was tested, the stabilizing sequence used to design each guide is indicated in the columns titled "5' Stabilizer SEQ ID NO." or "3' Stabilizer SEQ ID NO.". The unmodified guide sequences and chemically-modified guide sequences are provided in the sequence listing as indicated in the table below under the columns "Guide SEQ ID NO.", and "Mod. Guide SEQ ID NO." respectively. In this example, each table represents a separate experiment.

Table 29

Effect of guide stabilizing regions on % InDels of gene targets using Cas proteins and guides

Cas Protein		Guide					%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5' Stabilizing region SEQ ID NO.	3' Stabilizing region SEQ ID NO.	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V071	74	1859328	1172	N/A	778	952	34.74	2.46	DNMT1 Site 3
IONCAS-V071	74	1859342	N/A	N/A	766	940	12.15	0.27	
IONCAS-V071	74	1859692	488	N/A	779	953	21.73	3.49	
IONCAS-V071	74	1859693	489	N/A	780	954	41.44	3.51	
IONCAS-V071	74	1859704	491	N/A	781	955	17.22	1.84	
IONCAS-V071	74	1859708	490	N/A	782	956	27.30	1.86	
IONCAS-V071	74	1859712	493	N/A	783	957	65.55	2.50	
IONCAS-V071	74	1859716	492	N/A	784	958	45.66	2.48	
IONCAS-V071	74	1859329	1172	N/A	785	959	45.64	5.85	RUNX1 Site 4
IONCAS-V071	74	1859660	N/A	N/A	771	945	2.19	0.07	

IONCAS-V071	74	1859694	488	N/A	786	960	18.48	1.02	
IONCAS-V071	74	1859695	489	N/A	787	961	56.78	7.93	
IONCAS-V071	74	1859705	491	N/A	788	962	29.20	3.60	
IONCAS-V071	74	1859709	490	N/A	789	963	32.54	5.87	
IONCAS-V071	74	1859713	493	N/A	790	964	82.59	6.82	
IONCAS-V071	74	1859717	492	N/A	791	965	51.73	6.28	
AsCas12a	1	1859156	1175	N/A	792	966	64.97	13.13	RUNX1 Site 3
AsCas12a	1	1859337	N/A	N/A	793	967	28.60	1.72	
AsCas12a	1	1859696	488	N/A	794	968	88.66	21.07	
AsCas12a	1	1859697	489	N/A	795	969	93.85	34.33	
AsCas12a	1	1859706	491	N/A	796	970	85.07	11.08	
AsCas12a	1	1859710	490	N/A	797	971	85.54	10.44	
AsCas12a	1	1859714	493	N/A	798	972	88.54	4.48	
AsCas12a	1	1859718	492	N/A	799	973	82.39	15.09	
AsCas12a	1	1859327	1175	N/A	800	974	92.08	10.19	RUNX1 Site 4
AsCas12a	1	1859338	N/A	N/A	801	975	11.78	1.21	
AsCas12a	1	1859698	488	N/A	802	976	39.12	6.00	
AsCas12a	1	1859699	489	N/A	803	977	66.59	10.38	
AsCas12a	1	1859707	491	N/A	804	978	34.44	7.29	
AsCas12a	1	1859711	490	N/A	805	979	31.99	2.52	
AsCas12a	1	1859715	493	N/A	806	980	86.55	8.86	
AsCas12a	1	1859719	492	N/A	807	981	69.54	2.23	

Table 30

Effect of guide stabilizing regions on % InDels of gene targets using Cas proteins and guides

Cas Protein		Guide					%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5' Stabilizing region SEQ ID NO.	3' Stabilizing region SEQ ID NO.	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	37 ng	
IONCAS-V071	74	1859328	1172	N/A	778	952	33.85	2.92	DNMT1 Site 3
IONCAS-V071	74	1859342	N/A	N/A	766	940	15.63	0.36	
IONCAS-V071	74	1859693	489	N/A	780	954	53.38	8.52	
IONCAS-V071	74	1859712	493	N/A	783	957	42.94	2.96	
IONCAS-V071	74	1859700	N/A	494	808	982	22.09	1.50	
IONCAS-V071	74	1859329	1172	N/A	785	959	11.23	1.14	RUNX1 Site 4
IONCAS-V071	74	1859660	N/A	N/A	771	945	0.81	0.43	
IONCAS-V071	74	1859695	489	N/A	787	961	46.70	5.30	
IONCAS-V071	74	1859713	493	N/A	790	964	60.40	2.19	
IONCAS-V071	74	1859701	N/A	494	809	983	25.22	0.86	
AsCas12a	1	1859156	1175	N/A	792	966	32.01	4.71	
AsCas12a	1	1859337	N/A	N/A	793	967	19.98	6.04	

AsCas12a	1	1859697	489	N/A	795	969	86.86	27.99	RUNX 1 Site 3
AsCas12a	1	1859714	493	N/A	798	972	72.25	5.09	
AsCas12a	1	1859702	N/A	494	810	984	59.61	6.66	
AsCas12a	1	1859327	1175	N/A	800	974	23.17	2.93	RUNX 1 Site 4
AsCas12a	1	1859338	N/A	N/A	801	975	21.21	1.84	
AsCas12a	1	1859699	489	N/A	803	977	69.16	24.09	
AsCas12a	1	1859715	493	N/A	806	980	80.33	4.67	
AsCas12a	1	1859703	N/A	494	811	985	81.60	2.71	

Guide designs with stabilizing regions were selected from the screening experiments above and were further evaluated for dose-dependent effect on InDels of targets in HepG2 cells. Briefly, human HepG2 cells were split at the density of 19,000 cells per well in 96-well plates and cultured as described above. After 24 hours, culture media were changed.- RNA/LNP complexes containing Cas protein mRNA and guide RNA were generated, as described herein above, and added to the culture medium at total mRNA + guide RNA doses indicated in the tables below. Cells were harvested for InDel analysis after 72 hours, and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below. The half maximal effective concentration (EC_{50}) of each Cas protein and guide system was calculated using GraphPad Prism 10 software (GraphPad Software, San Diego, CA).

Table 31

Dose-dependent effect on InDels of RUNX1 Site 4 using IONCAS-V071 and guides with stabilizing regions

Cas Protein		Guide			%InDels						EC_{50} (ng)
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	990 ng	330 ng	110 ng	36.67 ng	12.22 ng	4.07 ng	
IONCAS-V071	74	1859329	785	959	13.13	11.23	3.03	1.14	0.21	0.19	180.4
IONCAS-V071	74	1859660	771	945	1.21	0.81	1.25	0.43	0.08	0.03	38.3
IONCAS-V071	74	1859695	787	961	44.66	46.70	17.04	5.30	1.68	0.53	116.3
IONCAS-V071	74	1859713	790	964	85.31	60.40	7.42	2.19	0.55	0.13	251.3

Table 32

Dose-dependent effect on InDels of RUNX1 Site 3 using AsCas12a and guides with stabilizing regions

Cas Protein		Guide			%InDels						EC_{50} (ng)
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	990 ng	330 ng	110 ng	36.67 ng	12.22 ng	4.07 ng	
AsCas12a	1	1859156	792	966	31.30	32.01	13.85	4.71	0.89	0.31	123.8

AsCas12a	1	1859337	793	967	14.22	19.98	11.09	6.04	0.64	0.10	58.1
AsCas12a	1	1859697	795	969	91.09	86.86	48.15	27.99	3.70	0.53	92.2
AsCas12a	1	1859714	798	972	95.76	72.25	9.79	5.09	0.09	0.18	234.6

Example 8: Design and effect of Cas proteins and guides in HepG2 cells, LNP delivery

Variants of Cas proteins selected from the examples above were designed by making mutations in one to eight residues in the Cas protein amino acid sequence. The Cas proteins were then screened for enzymatic activity in an *in vitro* assay wherein HepG2 cells were incubated with an LNP formulation of a Cas protein and guide editing system.

The Cas protein sequences (SEQ ID NO: 501-533, 600, 601) and GeneArt (ThermoFisher) human codon-optimized nucleic acid sequences (SEQ ID NO: 534-566, 603, 604) are provided in the sequence listing, as indicated in the table below. The Cas protein DNA sequences were cloned into a pTwistCMVpuro vector. To facilitate enzyme localization to nucleus, a nucleoplasmin-Nuclear Localization Signal (NLS) and two repeats of c-Myc NLS were added to the C-terminus of each Cas protein sequence. In addition, a HiBiT (Promega) tag was added between the nucleoplasmin NLS sequence and the first c-Myc NLS sequence for expression detection. These additional elements were linked to the proteins and translated to codon-optimized DNA, as indicated in the DNA sequences represented by SEQ ID NOs: 567-599, 606, 607 below.

Table 33

Cas protein variant sequences

Cas Protein ID	Cas Protein SEQ ID NO.	Cas protein codon-optimized DNA SEQ ID NO.	Cas protein + NLS + HiBit codon-optimized DNA SEQ ID NO.	Parent Cas Protein ID
IONCAS-V035en	501	534	567	IONCAS-V035
IONCAS-V035Plus	502	535	568	
IONCAS-V035hyper3aa	503	536	569	
IONCAS-V035ultra	504	537	570	
IONCAS-V035hyper4aa	505	538	571	
IONCAS-V035Super	506	539	572	
IONCAS-V035Super5aa	507	540	573	
IONCAS-V035enHiFi	600	603	606	
IONCAS-V050en	508	541	574	IONCAS-V050
IONCAS-V050Plus	509	542	575	
IONCAS-V050hyper3aa	510	543	576	
IONCAS-V050ultra	511	544	577	
IONCAS-V050hyper4aa	512	545	578	
IONCAS-V051en	513	546	579	IONCAS-V051

IONCAS-V051Plus	514	547	580	
IONCAS-V051hyper3aa	515	548	581	
IONCAS-V051ultra	516	549	582	
IONCAS-V051hyper4aa	517	550	583	
IONCAS-V051Super	518	551	584	
IONCAS-V051Super5aa	519	552	585	
IONCAS-V051enHiFi	601	604	607	
IONCAS-V071en	520	553	586	IONCAS-V071
IONCAS-V071Plus	521	554	587	
IONCAS-V071hyper3aa	522	555	588	
IONCAS-V071ultra	523	556	589	
IONCAS-V071hyper4aa	524	557	590	
IONCAS-V071Super	525	558	591	
IONCAS-V071Super5aa	526	559	592	
IONCAS-V071Uber	527	560	593	
IONCAS-V071Uber7aa	528	561	594	
IONCAS-V082en	529	562	595	IONCAS-V082
IONCAS-V082Plus	530	563	596	
IONCAS-V082hyper3aa	531	564	597	
IONCAS-V082ultra	532	565	598	
IONCAS-V082hyper4aa	533	566	599	

Guides were designed and synthesized comprising a 35 or 36 nucleotide protein-recognition region; a 23 nucleotide target-recognition region; and a three nucleotide UUU 3'-extension. Guides designed for IONCAS-V035 and IONCAS-V082 share a protein-recognition sequence. The sequence of each guide is provided in the sequence listing according to the SEQ ID NO. indicated in the tables below. Each guide in the tables below has the following modifications: the nucleosides at positions 1 and 2 at the 5'-end are 2'-OMe sugar moieties wherein each nucleoside is linked to the next nucleoside with a phosphorothioate internucleoside linkage; and the last two nucleosides at the 3'-end are 2'-OMe sugar moieties, wherein each nucleoside is linked to the preceding nucleoside by a phosphorothioate internucleoside linkage.

Human HepG2 cells were split at the density of 18,000 cells per well in 96-well plates and cultured in DMEM medium with 10% FBS and 1% PenStrep. After 24 hours, culture media were changed. RNA/LNP complexes were generated as described herein above. After incubating for 15 minutes, the mixture was added to the culture medium at total Cas protein mRNA + guide RNA doses indicated in the tables below. Cells were harvested for InDel analysis after 72 hours, and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below. Each table represents a separate experiment.

In the tables below, results are grouped by parent Cas protein, as indicated in the table titles. Results are further grouped by target, as indicated in the "Target" column. For each Cas protein and guide editing system that was tested, the full Cas protein amino acid sequence, guide sequence, and chemically-modified guide sequence are provided in the sequence listing as indicated in the table below under the columns "Cas Protein SEQ ID NO.", "Guide SEQ ID NO.", and "Mod. Guide SEQ ID NO." respectively.

Table 34

Effect of IONCAS-V035 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V035	38	1859332	812	986	20.50	0.51	FANCF Site 1
IONCAS-V035en	501	1859332	812	986	65.09	1.45	
IONCAS-V035Hyper3aa	503	1859332	812	986	61.16	1.19	
IONCAS-V035Plus	502	1859332	812	986	13.05	0.19	
IONCAS-V035Ultra	504	1859332	812	986	12.39	0.27	
IONCAS-V035	38	1859336	813	987	47.31	0.84	NLRC4
IONCAS-V035en	501	1859336	813	987	44.37	0.54	
IONCAS-V035Hyper3aa	503	1859336	813	987	49.62	0.81	
IONCAS-V035Plus	502	1859336	813	987	0.22	0.16	
IONCAS-V035Ultra	504	1859336	813	987	42.45	0.33	

Table 35

Effect of IONCAS-V035 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V035	38	1859332	812	986	53.61	11.32	FANCF Site 1
IONCAS-V035en	501	1859332	812	986	96.26	41.57	
IONCAS-V035Hyper3aa	503	1859332	812	986	97.44	40.41	
IONCAS-V035	38	1859336	813	987	91.36	24.78	NLRC4
IONCAS-V035en	501	1859336	813	987	91.50	27.88	
IONCAS-V035Hyper3aa	503	1859336	813	987	91.90	30.21	

Table 36

Effect of IONCAS-V050 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V050	53	1859334	814	988	6.87	0.11	FANCF Site 1
IONCAS-V050en	508	1859334	814	988	0.92	0.12	
IONCAS-V050Hyper3aa	510	1859334	814	988	39.77	0.49	
IONCAS-V050Plus	509	1859334	814	988	0.08	0.34	
IONCAS-V050Ultra	511	1859334	814	988	6.42	0.11	
IONCAS-V050	53	1859335	815	989	0.21	0.13	RUNX1 Site 3
IONCAS-V050en	508	1859335	815	989	0.18	0.06	
IONCAS-V050Hyper3aa	510	1859335	815	989	9.53	0.11	
IONCAS-V050Plus	509	1859335	815	989	0.59	0.08	
IONCAS-V050Ultra	511	1859335	815	989	0.12	0.06	

Table 37

Effect of IONCAS-V050 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V050	53	1859334	814	988	43.36	4.62	FANCF Site 1
IONCAS-V050Hyper3aa	510	1859334	814	988	93.73	34.94	
IONCAS-V050	53	1859335	815	989	1.24	0.38	RUNX1 Site 3
IONCAS-V050Hyper3aa	510	1859335	815	989	73.90	34.64	

Table 38

Effect of IONCAS-V051 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V051	54	1859330	816	990	4.46	0.18	FANCF Site 1
IONCAS-V051en	513	1859330	816	990	40.00	0.67	

IONCAS-V051Hyper3aa	515	1859330	816	990	9.56	0.19	
IONCAS-V051Plus	514	1859330	816	990	1.67	0.16	
IONCAS-V051Ultra	516	1859330	816	990	3.14	0.12	
IONCAS-V051	54	1859331	817	991	38.91	0.74	NLRC4
IONCAS-V051en	513	1859331	817	991	54.57	0.91	
IONCAS-V051Hyper3aa	515	1859331	817	991	49.82	0.72	
IONCAS-V051Plus	514	1859331	817	991	7.68	0.16	
IONCAS-V051Ultra	516	1859331	817	991	30.66	0.23	

Table 39

Effect of IONCAS-V051 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V051	54	1859330	816	990	14.72	2.88	FANCF Site 1
IONCAS-V051en	513	1859330	816	990	94.01	23.73	
IONCAS-V051Hyper3aa	515	1859330	816	990	43.23	4.77	
IONCAS-V051	54	1859331	817	991	89.18	27.18	NLRC4
IONCAS-V051en	513	1859331	817	991	90.09	31.12	
IONCAS-V051Hyper3aa	515	1859331	817	991	90.14	38.79	

Table 40

Effect of IONCAS-V071 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V071	74	1859328	778	952	8.31	0.12	DNMT1 Site 3
IONCAS-V071en	520	1859328	778	952	24.03	0.09	
IONCAS-V071Hyper3aa	522	1859328	778	952	28.95	0.19	
IONCAS-V071Plus	521	1859328	778	952	0.06	0.03	
IONCAS-V071Ultra	523	1859328	778	952	31.72	0.22	
IONCAS-V071	74	1859329	785	959	22.03	0.38	RUNX1 Site 4
IONCAS-V071en	520	1859329	785	959	64.85	2.01	
IONCAS-V071Hyper3aa	522	1859329	785	959	64.12	1.35	
IONCAS-V071Plus	521	1859329	785	959	0.21	0.11	
IONCAS-V071Ultra	523	1859329	785	959	50.18	0.61	

Table 41

Effect of IONCAS-V071 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	**Cas Protein SEQ ID NO.**	**Guide ID**	**Guide SEQ ID NO.**	**Mod. Guide SEQ ID NO.**	**330 ng**	**33 ng**	
IONCAS-V071	74	1859328	778	952	76.32	17.91	DNMT1 Site 3
IONCAS-V071en	520	1859328	778	952	93.23	32.45	
IONCAS-V071Hyper3aa	522	1859328	778	952	95.13	47.88	
IONCAS-V071Ultra	523	1859328	778	952	96.63	56.74	
IONCAS-V071	74	1859329	785	959	81.57	37.63	RUNX1 Site 4
IONCAS-V071en	520	1859329	785	959	98.55	68.53	
IONCAS-V071Hyper3aa	522	1859329	785	959	97.71	71.95	
IONCAS-V071Ultra	523	1859329	785	959	97.36	58.43	

Table 42

Effect of IONCAS-V082 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	**Cas Protein SEQ ID NO.**	**Guide ID**	**Guide SEQ ID NO.**	**Mod. Guide SEQ ID NO.**	**330 ng**	**33 ng**	
IONCAS-V082	84	1859332	812	986	18.10	0.25	FANCF Site 1
IONCAS-V082en	529	1859332	812	986	7.87	0.17	
IONCAS-V082Hyper3aa	531	1859332	812	986	53.84	0.56	
IONCAS-V082Plus	530	1859332	812	986	22.25	0.19	
IONCAS-V082Ultra	532	1859332	812	986	7.09	0.10	
IONCAS-V082	84	1859333	818	992	0.11	0.10	AGBL1 matched site 1
IONCAS-V082en	529	1859333	818	992	0.10	0.13	
IONCAS-V082Hyper3aa	531	1859333	818	992	0.84	0.11	
IONCAS-V082Plus	530	1859333	818	992	0.22	0.06	
IONCAS-V082Ultra	532	1859333	818	992	0.10	0.06	

Table 43

Effect of IONCAS-V082 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V082	84	1859332	812	986	73.86	11.44	FANCF Site 1
IONCAS-V082Hyper3aa	531	1859332	812	986	98.28	55.39	
IONCAS-V082Plus	530	1859332	812	986	83.28	12.23	
IONCAS-V082	84	1859333	818	992	0.39	0.18	AGBL1 matched site 1
IONCAS-V082Hyper3aa	531	1859333	818	992	16.67	3.32	
IONCAS-V082Plus	530	1859333	818	992	1.11	0.09	

Human HepG2 cells were split at the density of 17,000 cells per well in 96-well plates and cultured as described herein above. RNA/LNP complexes were generated as described herein above. After incubating for 15 minutes, the mixture was added to the culture medium at total Cas protein mRNA + guide RNA doses indicated in the tables below. After 24 incubation, cell media was changed. Cells were harvested for InDel analysis after an additional 48-72 hours, and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below.

In the tables below, results are grouped by parent Cas protein, as indicated in the table titles. Results are further grouped by target, as indicated in the "Target" column. For each Cas protein and guide editing system that was tested, the full Cas protein amino acid sequence, guide sequence, and chemically-modified guide sequence are provided in the sequence listing as indicated in the table below under the columns "Cas Protein SEQ ID NO.", "Guide SEQ ID NO.", and "Mod. Guide SEQ ID NO." respectively.

Table 44

Effect of IONCAS-V035 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	330 ng	33 ng	
IONCAS-V035	38	1859332	812	986	51.70	20.90	FANCF Site 1
IONCAS-V035en	501	1859332	812	986	93.25	80.40	
IONCAS-V035enHiFi	600	1859332	812	986	93.75	75.05	
IONCAS-V035hyper3aa	503	1859332	812	986	88.55	72.60	
IONCAS-V035hyper4aa	505	1859332	812	986	97.85	83.40	
IONCAS-V035Super	506	1859332	812	986	96.90	76.15	

IONCAS-V035Super5aa	507	1859332	812	986	96.20	78.45	
IONCAS-V035	38	1859336	813	987	91.50	48.85	NLRC4
IONCAS-V035en	501	1859336	813	987	93.60	48.05	
IONCAS-V035enHiFi	600	1859336	813	987	53.50	20.25	
IONCAS-V035hyper3aa	503	1859336	813	987	95.50	33.50	
IONCAS-V035hyper4aa	505	1859336	813	987	96.35	62.20	
IONCAS-V035Super	506	1859336	813	987	83.30	36.35	
IONCAS-V035Super5aa	507	1859336	813	987	80.65	50.15	

Table 45

Effect of IONCAS-V050 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	**SEQ ID NO.**	**Guide ID**	**Guide SEQ ID NO.**	**Mod. Guide SEQ ID NO.**	**330 ng**	**33 ng**	
IONCAS-V050	53	1859334	814	988	38.43	6.49	FANCF Site 1
IONCAS-V050hyper3aa	510	1859334	814	988	90.38	27.04	
IONCAS-V050hyper4aa	512	1859334	814	988	91.82	58.31	
IONCAS-V050	53	1859335	815	989	0.42	0.69	RUNX1 Site 3
IONCAS-V050hyper3aa	510	1859335	815	989	44.48	26.16	
IONCAS-V050hyper4aa	512	1859335	815	989	44.04	43.68	

Table 46

Effect of IONCAS-V051 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	**SEQ ID NO.**	**Guide ID**	**Guide SEQ ID NO.**	**Mod. Guide SEQ ID NO.**	**330 ng**	**33 ng**	
IONCAS-V051	54	1859330	816	990	20.25	9.27	FANCF Site 1
IONCAS-V051en	513	1859330	816	990	94.28	73.32	
IONCAS-V051enHiFi	601	1859330	816	990	91.26	65.11	
IONCAS-V051hyper3aa	515	1859330	816	990	41.46	16.30	
IONCAS-V051hyper4aa	517	1859330	816	990	31.63	14.77	
IONCAS-V051Super	518	1859330	816	990	95.72	69.33	
IONCAS-V051Super5aa	519	1859330	816	990	93.81	70.81	
IONCAS-V051	54	1859331	817	991	93.11	60.48	NLRC4
IONCAS-V051en	513	1859331	817	991	98.09	78.34	
IONCAS-V051enHiFi	601	1859331	817	991	92.23	59.93	
IONCAS-V051hyper3aa	515	1859331	817	991	96.16	70.11	

IONCAS-V051hyper4aa	517	1859331	817	991	95.63	72.39	
IONCAS-V051Super	518	1859331	817	991	97.84	77.80	
IONCAS-V051Super5aa	519	1859331	817	991	97.33	78.34	

Table 47

Effect of IONCAS-V071 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	**SEQ ID NO.**	**Guide ID**	**Guide SEQ ID NO.**	**Mod. Guide SEQ ID NO.**	**330 ng**	**33 ng**	
IONCAS-V071	74	1859328	778	952	60.65	15.04	DNMT1 Site 3
IONCAS-V071en	520	1859328	778	952	92.61	23.71	
IONCAS-V071Ultra	523	1859328	778	952	94.59	64.30	
IONCAS-V071Hyper3aa	522	1859328	778	952	91.69	47.83	
IONCAS-V071hyper4aa	524	1859328	778	952	92.57	37.65	
IONCAS-V071Super	525	1859328	778	952	75.83	27.30	
IONCAS-V071Super5aa	526	1859328	778	952	73.18	35.66	
IONCAS-V071Uber	527	1859328	778	952	45.24	14.89	
IONCAS-V071Uber7aa	528	1859328	778	952	47.07	18.05	
IONCAS-V071	74	1859329	785	959	68.11	45.48	RUNX1 Site 4
IONCAS-V071en	520	1859329	785	959	96.89	85.44	
IONCAS-V071Ultra	523	1859329	785	959	94.03	76.17	
IONCAS-V071Hyper3aa	522	1859329	785	959	93.81	73.61	
IONCAS-V071hyper4aa	524	1859329	785	959	95.39	78.29	
IONCAS-V071Super	525	1859329	785	959	93.61	74.63	
IONCAS-V071Super5aa	526	1859329	785	959	92.40	75.37	
IONCAS-V071Uber	527	1859329	785	959	80.81	50.30	
IONCAS-V071Uber7aa	528	1859329	785	959	80.07	46.03	

Table 48

Effect of IONCAS-V082 variants and guides on % InDels of gene targets

Cas Protein		Guide			%InDels		Target
Cas Protein ID	**SEQ ID NO.**	**Guide ID**	**Guide SEQ ID NO.**	**Mod. Guide SEQ ID NO.**	**330 ng**	**33 ng**	
IONCAS-V082	84	1859332	812	986	67.70	25.58	FANCF Site 1
IONCAS-V082hyper3aa	531	1859332	812	986	97.90	63.73	
IONCAS-V082hyper4aa	533	1859332	812	986	98.58	85.57	
IONCAS-V082	84	1859333	818	992	0.50	0.16	

IONCAS-V082hyper3aa	531	1859333	818	992	13.29	2.63	AGBL1 matched site 1
IONCAS-V082hyper4aa	533	1859333	818	992	10.43	3.49	

Example 9: Effect of guide stabilizing regions on Cas protein activity in AML12 cells, LNP delivery

Modified guides comprising stabilizing regions appended at the 5'- or 3'-end were designed for AsCas12a. The guides were screened for their effect on the enzymatic activity of their respective Cas proteins in an *in vitro* assay wherein AML12 cells were incubated with an LNP formulation of a Cas protein and guide editing system.

Guides targeted to two sites on mouse TTR (designated G03 and G06) were designed comprising a protein-recognition region comprising a 20 nucleotide direct repeat region; a 20 nucleotide target-recognition region; and a stabilizing region selected from Table 2 appended to the 5'- or 3'-end by a UUUU linker sequence. Each guide was capped at the 3'-end with UUU. Guides that were designed without a 5'- stabilizing region were instead capped at the 5'-end with UU.

The sequence of each guide is provided in the sequence listing according to the SEQ ID NO. indicated in the tables below. Each guide in the tables below has the following modifications: the nucleosides at positions 1 and 2 at the 5'-end are 2'-OMe sugar moieties wherein each nucleoside is linked to the next nucleoside with phosphorothioate internucleoside linkage; and the last two nucleosides at the 3'-end are 2'-OMe sugar moieties, wherein each nucleoside is linked to the preceding nucleoside by a phosphorothioate internucleoside linkage.

Mouse AML12 cells were split at the density of 16,000 cells per well in 96-well plates and cultured in DMEM F12 (Gibco) medium with 10% FBS, 1% ITS, 1% PenStrep, and 0.001% dexamethasone. RNA/LNP complexes were generated by mixing RNA with a LNP formulation, as described herein above. After incubating for 15 minutes, the mixture was added to the culture medium at total Cas protein mRNA + guide RNA doses indicated in the tables below. After 24 hours, the cell media was changed, and the cells were incubated for an additional 72 hours. Afterwards, cells were harvested and InDel analysis was performed as described herein above. Results are presented as Percentage of InDels (%InDels) in the table below.

In the tables below, results are grouped by Cas protein, as indicated in the "Cas Protein ID" column, and by target, as indicated in the "Target" column. For each Cas protein and guide editing system that was tested, the stabilizing sequence used to design each guide is indicated in the columns titled "5' Stabilizer SEQ ID NO." or "3' Stabilizer SEQ ID NO.". The unmodified guide sequences and chemically-modified guide sequences are provided in the sequence listing as indicated in the table below under the columns "Guide SEQ

ID NO.", and "Mod. Guide SEQ ID NO." respectively. In this example, each table represents a separate experiment.

Table 49

Effect of guide stabilizing regions on % InDels of gene targets using IONCAS-V035

Cas Protein		Guide					%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5' Stabilizer SEQ ID NO.	3' Stabilizer SEQ ID NO.	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	660 ng	220 ng	
IONCAS-V035	38	1887453	N/A	N/A	1013	1047	0.08	0.14	mTTR G03
IONCAS-V035	38	1887483	498	N/A	1014	1048	0.16	0.37	
IONCAS-V035	38	1887014	493	N/A	1015	1049	0.15	0.08	
IONCAS-V035	38	1887015	N/A	498	1016	1050	0.00	0.03	
IONCAS-V035	38	1887016	N/A	493	1017	1051	0.09	0.09	
IONCAS-V035	38	1887002	498	493	1018	1052	0.10	0.28	
IONCAS-V035	38	1887005	493	498	1019	1053	0.05	0.16	
IONCAS-V035	38	1887083	N/A	N/A	1020	1054	0.90	0.96	mTTR G06
IONCAS-V035	38	1887113	498	N/A	1021	1055	28.09	25.08	
IONCAS-V035	38	1887018	493	N/A	1022	1056	2.17	3.47	
IONCAS-V035	38	1887019	N/A	498	1023	1057	0.70	0.67	
IONCAS-V035	38	1887020	N/A	493	1024	1058	1.63	1.97	
IONCAS-V035	38	1887008	498	493	1025	1059	2.18	2.03	
IONCAS-V035	38	1887010	493	498	1026	1060	0.68	0.50	

5

Table 50

Effect of guide stabilizing regions on % InDels of gene targets using Cas proteins and guides

Cas Protein		Guide					%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5' Stabilizer SEQ ID NO.	3' Stabilizer SEQ ID NO.	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	660 ng	220 ng	
IONCAS-V071	74	1887453	N/A	N/A	1013	1047	0.10	0.08	mTTR G03
IONCAS-V071	74	1887483	498	N/A	1014	1048	0.19	0.39	
IONCAS-V071	74	1887014	493	N/A	1015	1049	0.10	0.23	
IONCAS-V071	74	1887015	N/A	498	1016	1050	0.00	0.07	
IONCAS-V071	74	1887016	N/A	493	1017	1051	0.11	0.06	
IONCAS-V071	74	1887002	498	493	1018	1052	0.09	0.17	
IONCAS-V071	74	1887005	493	498	1019	1053	0.09	0.06	
IONCAS-V071	74	1887083	N/A	N/A	1020	1054	0.40	0.22	mTTR G06
IONCAS-V071	74	1887113	498	N/A	1021	1055	4.74	4.21	

IONCAS-V071	74	1887018	493	N/A	1022	1056	1.10	0.88	
IONCAS-V071	74	1887019	N/A	498	1023	1057	0.59	0.40	
IONCAS-V071	74	1887020	N/A	493	1024	1058	0.94	0.63	
IONCAS-V071	74	1887008	498	493	1025	1059	1.60	0.35	
IONCAS-V071	74	1887010	493	498	1026	1060	0.68	0.45	

Table 51

Effect of guide stabilizing regions on % InDels of gene targets using Cas proteins and guides

Cas Protein		Guide					%InDels		Target
Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	5' Stabilizer SEQ ID NO.	3' Stabilizer SEQ ID NO.	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	660 ng	220 ng	
AsCas12a	1	1887453	N/A	N/A	1013	1047	0.05	0.67	mTTR G03
AsCas12a	1	1887483	498	N/A	1014	1048	14.54	6.18	
AsCas12a	1	1887014	493	N/A	1015	1049	2.31	2.22	
AsCas12a	1	1887015	N/A	498	1016	1050	0.82	0.26	
AsCas12a	1	1887016	N/A	493	1017	1051	0.47	0.47	
AsCas12a	1	1887002	498	493	1018	1052	3.96	1.79	
AsCas12a	1	1887005	493	498	1019	1053	0.58	0.95	
AsCas12a	1	1887083	N/A	N/A	1020	1054	1.83	2.21	mTTR G06
AsCas12a	1	1887113	498	N/A	1021	1055	63.11	59.16	
AsCas12a	1	1887018	493	N/A	1022	1056	12.39	8.86	
AsCas12a	1	1887019	N/A	498	1023	1057	7.72	11.16	
AsCas12a	1	1887020	N/A	493	1024	1058	13.47	6.15	
AsCas12a	1	1887008	498	493	1025	1059	13.83	11.59	
AsCas12a	1	1887010	493	498	1026	1060	6.59	3.23	

Example 10: Cas editing system sequences provided in the sequence listing

For each Cas editing system, the Cas protein ID and its SEQ ID NO. are indicated in the columns under "Cas Protein", and the Guide ID and its SEQ ID NO. are indicated in the columns under "Guide." The sequences of protein-recognition regions and direct repeat regions of guides are included in the sequence listing as indicated in the table below. The following sequences can be found in the sequence listing as indicated by their respective SEQ ID NOs: complete guide sequences are provided as indicated in the column "Guide SEQ ID NO."; chemically-modified complete guide sequences are provided as indicated in the column "Mod. Guide SEQ ID NO."; direct repeat region sequences are provided as indicated in the column "Direct repeat region SEQ ID NO."; and protein-recognition region sequences are provided as indicated in the column "Protein-recognition region SEQ ID NO." "N/A" indicates that a guide is not chemically-modified.

Table 52

SEQ ID NOs of guides and their respective protein-recognition and direct repeat regions

Cas Protein ID	Cas Protein SEQ ID NO.	Guide ID	Guide SEQ ID NO.	Mod. Guide SEQ ID NO.	Protein-Recognition Region SEQ ID NO.	Direct Repeat Region
AsCas12a	1	AsCas12a gRNA	465	911	1152	1228
AsCas12a	1	gAS-DNM-1	625	N/A	1152	1228
AsCas12a	1	gAS-DNM-3	640	N/A	1152	1228
AsCas12a	1	gAS-FA-1	653	N/A	1152	1228
AsCas12a	1	gAS-FA-2	666	N/A	1152	1228
AsCas12a	1	gAS-NLR	680	N/A	1152	1228
AsCas12a	1	gAS-MS-1	693	N/A	1152	1228
AsCas12a	1	gAS-RUN-2	708	N/A	1152	1228
AsCas12a	1	gAS-RUN-3	723	N/A	1152	1228
AsCas12a	1	gAS-RUN-4	737	N/A	1152	1228
AsCas12a	1	1859156	792	966	1152	1228
AsCas12a	1	1859337	793	967	1153	1228
AsCas12a	1	1859327	800	974	1152	1228
AsCas12a	1	1859338	801	975	1153	1228
AsCas12a	1	1859696	794	968	1162	1228
AsCas12a	1	1859697	795	969	1163	1228
AsCas12a	1	1859706	796	970	1164	1228
AsCas12a	1	1859710	797	971	1165	1228
AsCas12a	1	1859714	798	972	1166	1228
AsCas12a	1	1859718	799	973	1167	1228
AsCas12a	1	1859698	802	976	1162	1228
AsCas12a	1	1859699	803	977	1163	1228
AsCas12a	1	1859707	804	978	1164	1228
AsCas12a	1	1859711	805	979	1165	1228
AsCas12a	1	1859715	806	980	1166	1228
AsCas12a	1	1859719	807	981	1167	1228
AsCas12a	1	1859702	810	984	1153	1228
AsCas12a	1	1859703	811	985	1153	1228
AsCas12a	1	1887453	1013	1047	1170	1228
AsCas12a	1	1887483	1014	1048	1171	1228
AsCas12a	1	1887014	1015	1049	1166	1228
AsCas12a	1	1887015	1016	1050	1170	1228
AsCas12a	1	1887016	1017	1051	1170	1228
AsCas12a	1	1887002	1018	1052	1171	1228
AsCas12a	1	1887005	1019	1053	1166	1228
AsCas12a	1	1887083	1020	1054	1170	1228
AsCas12a	1	1887113	1021	1055	1171	1228

AsCas12a	1	1887018	1022	1056	1166	1228
AsCas12a	1	1887019	1023	1057	1170	1228
AsCas12a	1	1887020	1024	1058	1170	1228
AsCas12a	1	1887008	1025	1059	1171	1228
AsCas12a	1	1887010	1026	1060	1166	1228
IONCAS-V001	4	V1gRNA	373	819	1061	1176
IONCAS-V002	5	V2gRNA	374	820	1062	1177
IONCAS-V003	6	V3gRNA	375	821	1063	1178
IONCAS-V004	7	V4gRNA	376	822	1064	1179
IONCAS-V005	8	V5gRNA	377	823	1065	1180
IONCAS-V006	9	V6gRNA	378	824	1066	1181
IONCAS-V007	10	V7gRNA	379	825	1067	1179
IONCAS-V008	11	V8gRNA	380	826	1068	1182
IONCAS-V009	12	V9gRNA	381	827	1069	1183
IONCAS-V010	13	V10gRNA	382	828	1070	1184
IONCAS-V011	14	V11gRNA	383	829	1071	1185
IONCAS-V012	15	V12gRNA	384	830	1072	1186
IONCAS-V013	16	V13gRNA	385	831	1073	1179
IONCAS-V014	17	V14gRNA	386	832	1074	1187
IONCAS-V015	18	V15gRNA	387	833	1075	1188
IONCAS-V016	19	V16gRNA	388	834	1076	1178
IONCAS-V017	20	V17gRNA	389	835	1077	1180
IONCAS-V018	21	V18gRNA	390	836	1078	1189
IONCAS-V019	22	V19gRNA	391	837	1079	1185
IONCAS-V020	23	V20gRNA	392	838	1080	1178
IONCAS-V021	24	V21gRNA	393	839	1081	1178
IONCAS-V021	24	gV21-DNM-1	611	N/A	1081	1178
IONCAS-V021	24	gV21-DNM-3	626	N/A	1081	1178
IONCAS-V021	24	gV21-FA-1	641	N/A	1081	1178
IONCAS-V021	24	gV21-FA-2	654	N/A	1081	1178
IONCAS-V021	24	gV21-NLR	667	N/A	1081	1178
IONCAS-V021	24	gV21-MS-1	681	N/A	1081	1178
IONCAS-V021	24	gV21-RUN-2	694	N/A	1081	1178
IONCAS-V021	24	gV21-RUN-3	709	N/A	1081	1178
IONCAS-V021	24	gV21-RUN-4	724	N/A	1081	1178
IONCAS-V022	25	V22gRNA	394	840	1082	1190
IONCAS-V023	26	V23gRNA	395	841	1083	1191
IONCAS-V023	26	V23gAS-DNM-1	625	N/A	1152	1228
IONCAS-V023	26	V23gAS-DNM-3	640	N/A	1152	1228
IONCAS-V023	26	V23gAS-FA-1	653	N/A	1152	1228
IONCAS-V023	26	V23gAS-FA-2	666	N/A	1152	1228
IONCAS-V023	26	V23gAS-NLR	680	N/A	1152	1228
IONCAS-V023	26	V23gAS-MS-1	693	N/A	1152	1228

IONCAS-V023	26	V23gAS-RUN-2	708	N/A	1152	1228
IONCAS-V023	26	V23gAS-RUN-3	723	N/A	1152	1228
IONCAS-V023	26	V23gAS-RUN-4	737	N/A	1152	1228
IONCAS-V024	27	V24gRNA	396	842	1084	1180
IONCAS-V024	27	gV24-DNM-1	612	N/A	1084	1180
IONCAS-V024	27	gV24-DNM-3	627	N/A	1084	1180
IONCAS-V024	27	gV24-NLR	668	N/A	1084	1180
IONCAS-V024	27	gV24-MS-1	682	N/A	1084	1180
IONCAS-V024	27	gV24-RUN-2	695	N/A	1084	1180
IONCAS-V024	27	gV24-RUN-3	710	N/A	1084	1180
IONCAS-V024	27	gV24-RUN-4	725	N/A	1084	1180
IONCAS-V025	28	V25gRNA	397	843	1085	1178
IONCAS-V025	28	gV25-DNM-1	613	N/A	1085	1178
IONCAS-V025	28	gV25-DNM-3	628	N/A	1085	1178
IONCAS-V025	28	gV25-FA-1	642	N/A	1085	1178
IONCAS-V025	28	gV25-FA-2	655	N/A	1085	1178
IONCAS-V025	28	gV25-NLR	669	N/A	1085	1178
IONCAS-V025	28	gV25-MS-1	683	N/A	1085	1178
IONCAS-V025	28	gV25-RUN-2	696	N/A	1085	1178
IONCAS-V025	28	gV25-RUN-3	711	N/A	1085	1178
IONCAS-V025	28	gV25-RUN-4	726	N/A	1085	1178
IONCAS-V026	29	V26gRNA	398	844	1086	1180
IONCAS-V026	29	gV26-DNM-1	614	N/A	1086	1180
IONCAS-V026	29	gV26-DNM-3	629	N/A	1086	1180
IONCAS-V026	29	gV26-NLR	670	N/A	1086	1180
IONCAS-V026	29	gV26-MS-1	684	N/A	1086	1180
IONCAS-V026	29	gV26-RUN-2	697	N/A	1086	1180
IONCAS-V026	29	gV26-RUN-3	712	N/A	1086	1180
IONCAS-V026	29	gV26-RUN-4	727	N/A	1086	1180
IONCAS-V027	30	V27gRNA	399	845	1087	1192
IONCAS-V027	30	gV27-DNM-1	615	N/A	1087	1192
IONCAS-V027	30	gV27-DNM-3	630	N/A	1087	1192
IONCAS-V027	30	gV27-FA-1	643	N/A	1087	1192
IONCAS-V027	30	gV27-FA-2	656	N/A	1087	1192
IONCAS-V027	30	gV27-NLR	671	N/A	1087	1192
IONCAS-V027	30	gV27-MS-1	685	N/A	1087	1192
IONCAS-V027	30	gV27-RUN-2	698	N/A	1087	1192
IONCAS-V027	30	gV27-RUN-3-4	713	N/A	1087	1192
IONCAS-V028	31	V28gRNA	400	846	1088	1193
IONCAS-V029	32	V29gRNA	401	847	1089	1178
IONCAS-V029	32	gV29-DNM-1	616	N/A	1089	1178
IONCAS-V029	32	gV29-DNM-3	631	N/A	1089	1178
IONCAS-V029	32	gV29-FA-1	644	N/A	1089	1178

IONCAS-V029	32	gV29-FA-2	657	N/A	1089	1178
IONCAS-V029	32	gV29-NLR	672	N/A	1089	1178
IONCAS-V029	32	gV29-MS-1	686	N/A	1089	1178
IONCAS-V029	32	gV29-RUN-2	699	N/A	1089	1178
IONCAS-V029	32	gV29-RUN-3	714	N/A	1089	1178
IONCAS-V029	32	gV29-RUN-4	728	N/A	1089	1178
IONCAS-V030	33	V30gRNA	402	848	1090	1178
IONCAS-V031	34	V31gRNA	403	849	1091	1194
IONCAS-V031	34	V31gAS-DNM-1	625	N/A	1152	1228
IONCAS-V031	34	V31gAS-DNM-3	640	N/A	1152	1228
IONCAS-V031	34	V31gAS-FA-1	653	N/A	1152	1228
IONCAS-V031	34	V31gAS-FA-2	666	N/A	1152	1228
IONCAS-V031	34	V31gAS-NLR	680	N/A	1152	1228
IONCAS-V031	34	V31gAS-MS-1	693	N/A	1152	1228
IONCAS-V031	34	V31gAS-RUN-2	708	N/A	1152	1228
IONCAS-V031	34	V31gAS-RUN-3	723	N/A	1152	1228
IONCAS-V031	34	V31gAS-RUN-4	737	N/A	1152	1228
IONCAS-V032	35	V32gRNA	404	850	1092	1178
IONCAS-V033	36	V33gRNA	405	851	1093	1179
IONCAS-V034	37	V34gRNA	406	852	1094	1180
IONCAS-V034	37	gV34-DNM-1	617	N/A	1094	1180
IONCAS-V034	37	gV34-DNM-3	632	N/A	1094	1180
IONCAS-V034	37	gV34-FA-1	645	N/A	1094	1180
IONCAS-V034	37	gV34-FA-2	658	N/A	1094	1180
IONCAS-V034	37	gV34-NLR	673	N/A	1094	1180
IONCAS-V034	37	gV34-MS-1	687	N/A	1094	1180
IONCAS-V034	37	gV34-RUN-2	700	N/A	1094	1180
IONCAS-V034	37	gV34-RUN-3	715	N/A	1094	1180
IONCAS-V034	37	gV34-RUN-4	729	N/A	1094	1180
IONCAS-V035	38	V35gRNA	407	853	1095	1195
IONCAS-V035	38	V35gAS-DNM-1	625	N/A	1152	1228
IONCAS-V035	38	V35gAS-DNM-3	640	N/A	1152	1228
IONCAS-V035	38	V35gAS-FA-1	653	N/A	1152	1228
IONCAS-V035	38	V35gAS-FA-2	666	N/A	1152	1228
IONCAS-V035	38	V35gAS-NLR	680	N/A	1152	1228
IONCAS-V035	38	V35gAS-MS-1	693	N/A	1152	1228
IONCAS-V035	38	V35gAS-RUN-2	708	N/A	1152	1228
IONCAS-V035	38	V35gAS-RUN-3	723	N/A	1152	1228
IONCAS-V035	38	V35gAS-RUN-4	737	N/A	1152	1228
IONCAS-V035	38	1859672	738	912	1153	1228
IONCAS-V035	38	1859673	739	913	1153	1228
IONCAS-V035	38	1859674	740	914	1153	1228
IONCAS-V035	38	1859675	741	915	1153	1228

IONCAS-V035	38	1859676	742	916	1153	1228
IONCAS-V035	38	1859687	743	917	1153	1228
IONCAS-V035	38	1859688	744	918	1153	1228
IONCAS-V035	38	1859689	745	919	1153	1228
IONCAS-V035	38	1859690	746	920	1153	1228
IONCAS-V035	38	1859691	747	921	1153	1228
IONCAS-V035	38	1859332	812	986	1152	1228
IONCAS-V035	38	1859336	813	987	1152	1228
IONCAS-V035	38	1887453	1013	1047	1170	1228
IONCAS-V035	38	1887483	1014	1048	1171	1228
IONCAS-V035	38	1887014	1015	1049	1166	1228
IONCAS-V035	38	1887015	1016	1050	1170	1228
IONCAS-V035	38	1887016	1017	1051	1170	1228
IONCAS-V035	38	1887002	1018	1052	1171	1228
IONCAS-V035	38	1887005	1019	1053	1166	1228
IONCAS-V035	38	1887083	1020	1054	1170	1228
IONCAS-V035	38	1887113	1021	1055	1171	1228
IONCAS-V035	38	1887018	1022	1056	1166	1228
IONCAS-V035	38	1887019	1023	1057	1170	1228
IONCAS-V035	38	1887020	1024	1058	1170	1228
IONCAS-V035	38	1887008	1025	1059	1171	1228
IONCAS-V035	38	1887010	1026	1060	1166	1228
IONCAS-V035en	501	1859332	812	986	1152	1228
IONCAS-V035en	501	1859336	813	987	1152	1228
IONCAS-V035enHiFi	600	1859332	812	986	1152	1228
IONCAS-V035enHiFi	600	1859336	813	987	1152	1228
IONCAS-V035Hyper3aa	503	1859332	812	986	1152	1228
IONCAS-V035Hyper3aa	503	1859336	813	987	1152	1228
IONCAS-V035hyper4aa	505	1859332	812	986	1152	1228
IONCAS-V035hyper4aa	505	1859336	813	987	1152	1228
IONCAS-V035Plus	502	1859332	812	986	1152	1228
IONCAS-V035Plus	502	1859336	813	987	1152	1228
IONCAS-V035Super	506	1859332	812	986	1152	1228
IONCAS-V035Super	506	1859336	813	987	1152	1228
IONCAS-V035Super5aa	507	1859332	812	986	1152	1228
IONCAS-V035Super5aa	507	1859336	813	987	1152	1228
IONCAS-V035Ultra	504	1859332	812	986	1152	1228
IONCAS-V035Ultra	504	1859336	813	987	1152	1228
IONCAS-V036	39	V36gRNA	408	854	1096	1189
IONCAS-V037	40	V37gRNA	409	855	1097	1180
IONCAS-V038	41	V38gRNA	410	856	1098	1196
IONCAS-V039	42	V39gRNA	411	857	1099	1179
IONCAS-V039	42	gV39-DNM-1	618	N/A	1099	1179

IONCAS-V039	42	gV39-DNM-3	633	N/A	1099	1179
IONCAS-V039	42	gV39-FA-1	646	N/A	1099	1179
IONCAS-V039	42	gV39-FA-2	659	N/A	1099	1179
IONCAS-V039	42	gV39-NLR	674	N/A	1099	1179
IONCAS-V039	42	gV39-MS-1	688	N/A	1099	1179
IONCAS-V039	42	gV39-RUN-2	701	N/A	1099	1179
IONCAS-V039	42	gV39-RUN-3	716	N/A	1099	1179
IONCAS-V039	42	gV39-RUN-4	730	N/A	1099	1179
IONCAS-V040	43	V40gRNA	412	858	1100	1179
IONCAS-V041	44	V41gRNA	413	859	1101	1197
IONCAS-V042	45	V42gRNA	414	860	1102	1191
IONCAS-V042	45	gV42-DNM-1	619	N/A	1102	1191
IONCAS-V042	45	gV42-DNM-3	634	N/A	1102	1191
IONCAS-V042	45	gV42-FA-1	647	N/A	1102	1191
IONCAS-V042	45	gV42-FA-2	660	N/A	1102	1191
IONCAS-V042	45	gV42-NLR	675	N/A	1102	1191
IONCAS-V042	45	gV42-MS-1	689	N/A	1102	1191
IONCAS-V042	45	gV42-RUN-2	702	N/A	1102	1191
IONCAS-V042	45	gV42-RUN-3	717	N/A	1102	1191
IONCAS-V042	45	gV42-RUN-4	731	N/A	1102	1191
IONCAS-V043	46	V43gRNA	415	861	1103	1198
IONCAS-V043	46	V43gAS-DNM-1	625	N/A	1152	1228
IONCAS-V043	46	V43gAS-DNM-3	640	N/A	1152	1228
IONCAS-V043	46	V43gAS-FA-1	653	N/A	1152	1228
IONCAS-V043	46	V43gAS-FA-2	666	N/A	1152	1228
IONCAS-V043	46	V43gAS-NLR	680	N/A	1152	1228
IONCAS-V043	46	V43gAS-MS-1	693	N/A	1152	1228
IONCAS-V043	46	V43gAS-RUN-2	708	N/A	1152	1228
IONCAS-V043	46	V43gAS-RUN-3	723	N/A	1152	1228
IONCAS-V043	46	V43gAS-RUN-4	737	N/A	1152	1228
IONCAS-V044	47	V44gRNA	416	862	1104	1178
IONCAS-V044	47	gV44-DNM-1	620	N/A	1104	1178
IONCAS-V044	47	gV44-DNM-3	635	N/A	1104	1178
IONCAS-V044	47	gV44-FA-1	648	N/A	1104	1178
IONCAS-V044	47	gV44-FA-2	661	N/A	1104	1178
IONCAS-V044	47	gV44-NLR	676	N/A	1104	1178
IONCAS-V044	47	gV44-MS-1	690	N/A	1104	1178
IONCAS-V044	47	gV44-RUN-2	703	N/A	1104	1178
IONCAS-V044	47	gV44-RUN-3	718	N/A	1104	1178
IONCAS-V044	47	gV44-RUN-4	732	N/A	1104	1178
IONCAS-V045	48	V45gRNA	417	863	1105	1199
IONCAS-V046	49	V46gRNA	418	864	1106	1178
IONCAS-V047	50	V47gRNA	419	865	1107	1185

IONCAS-V047	50	V47gAS-DNM-1	625	N/A	1152	1228
IONCAS-V047	50	V47gAS-DNM-3	640	N/A	1152	1228
IONCAS-V047	50	V47gAS-FA-1	653	N/A	1152	1228
IONCAS-V047	50	V47gAS-FA-2	666	N/A	1152	1228
IONCAS-V047	50	V47gAS-NLR	680	N/A	1152	1228
IONCAS-V047	50	V47gAS-MS-1	693	N/A	1152	1228
IONCAS-V047	50	V47gAS-RUN-2	708	N/A	1152	1228
IONCAS-V047	50	V47gAS-RUN-3	723	N/A	1152	1228
IONCAS-V047	50	V47gAS-RUN-4	737	N/A	1152	1228
IONCAS-V048	51	V48gRNA	420	866	1108	1200
IONCAS-V049	52	V49gRNA	421	867	1109	1178
IONCAS-V050	53	V50gRNA	422	868	1110	1178
IONCAS-V050	53	gV50-DNM-1	621	N/A	1110	1178
IONCAS-V050	53	gV50-DNM-3	636	N/A	1110	1178
IONCAS-V050	53	gV50-FA-1	649	N/A	1110	1178
IONCAS-V050	53	gV50-FA-2	662	N/A	1110	1178
IONCAS-V050	53	gV50-RUN-2	704	N/A	1110	1178
IONCAS-V050	53	gV50-RUN-3	719	N/A	1110	1178
IONCAS-V050	53	gV50-RUN-4	733	N/A	1110	1178
IONCAS-V050	53	1859662	748	922	1154	1178
IONCAS-V050	53	1859663	749	923	1154	1178
IONCAS-V050	53	1859664	750	924	1154	1178
IONCAS-V050	53	1859665	751	925	1154	1178
IONCAS-V050	53	1859666	752	926	1154	1178
IONCAS-V050	53	1859682	753	927	1154	1178
IONCAS-V050	53	1859683	754	928	1154	1178
IONCAS-V050	53	1859684	755	929	1154	1178
IONCAS-V050	53	1859685	756	930	1154	1178
IONCAS-V050	53	1859686	757	931	1154	1178
IONCAS-V050	53	1859334	814	988	1110	1178
IONCAS-V050	53	1859662	748	922	1154	1178
IONCAS-V050	53	1859335	815	989	1110	1178
IONCAS-V050	53	1859334	814	988	1110	1178
IONCAS-V050en	508	1859334	814	988	1110	1178
IONCAS-V050en	508	1859335	815	989	1110	1178
IONCAS-V050Hyper3aa	510	1859334	814	988	1110	1178
IONCAS-V050Hyper3aa	510	1859335	815	989	1110	1178
IONCAS-V050hyper4aa	512	1859334	814	988	1110	1178
IONCAS-V050hyper4aa	512	1859335	815	989	1110	1178
IONCAS-V050Plus	509	1859334	814	988	1110	1178
IONCAS-V050Plus	509	1859335	815	989	1110	1178
IONCAS-V050Ultra	511	1859334	814	988	1110	1178
IONCAS-V050Ultra	511	1859335	815	989	1110	1178

IONCAS-V051	54	V51gRNA	423	869	1111	1178
IONCAS-V051	54	gV51-DNM-1	622	N/A	1111	1178
IONCAS-V051	54	gV51-DNM-3	637	N/A	1111	1178
IONCAS-V051	54	gV51-FA-1	650	N/A	1111	1178
IONCAS-V051	54	gV51-FA-2	663	N/A	1111	1178
IONCAS-V051	54	gV51-NLR	677	N/A	1111	1178
IONCAS-V051	54	gV51-RUN-2	705	N/A	1111	1178
IONCAS-V051	54	gV51-RUN-3	720	N/A	1111	1178
IONCAS-V051	54	gV51-RUN-4	734	N/A	1111	1178
IONCAS-V051	54	1859662	748	922	1154	1178
IONCAS-V051	54	1859663	749	923	1154	1178
IONCAS-V051	54	1859664	750	924	1154	1178
IONCAS-V051	54	1859665	751	925	1154	1178
IONCAS-V051	54	1859666	752	926	1154	1178
IONCAS-V051	54	1859667	758	932	1154	1178
IONCAS-V051	54	1859668	759	933	1154	1178
IONCAS-V051	54	1859669	760	934	1154	1178
IONCAS-V051	54	1859670	761	935	1154	1178
IONCAS-V051	54	1859671	762	936	1154	1178
IONCAS-V051	54	1886986	993	1027	1168	1178
IONCAS-V051	54	1886987	994	1028	1168	1178
IONCAS-V051	54	1886988	995	1029	1168	1178
IONCAS-V051	54	1886989	996	1030	1168	1178
IONCAS-V051	54	1886990	997	1031	1168	1178
IONCAS-V051	54	1886991	998	1032	1168	1178
IONCAS-V051	54	1886992	999	1033	1168	1178
IONCAS-V051	54	1886993	1000	1034	1168	1178
IONCAS-V051	54	1886994	1001	1035	1168	1178
IONCAS-V051	54	1886995	1002	1036	1168	1178
IONCAS-V051	54	1859330	816	990	1111	1178
IONCAS-V051	54	1859331	817	991	1111	1178
IONCAS-V051en	513	1859330	816	990	1111	1178
IONCAS-V051en	513	1859331	817	991	1111	1178
IONCAS-V051enHiFi	601	1859330	816	990	1111	1178
IONCAS-V051enHiFi	601	1859331	817	991	1111	1178
IONCAS-V051Hyper3aa	515	1859330	816	990	1111	1178
IONCAS-V051Hyper3aa	515	1859331	817	991	1111	1178
IONCAS-V051hyper4aa	517	1859330	816	990	1111	1178
IONCAS-V051hyper4aa	517	1859331	817	991	1111	1178
IONCAS-V051Plus	514	1859330	816	990	1111	1178
IONCAS-V051Plus	514	1859331	817	991	1111	1178
IONCAS-V051Super	518	1859330	816	990	1111	1178
IONCAS-V051Super	518	1859331	817	991	1111	1178

IONCAS-V051Super5aa	519	1859330	816	990	1111	1178
IONCAS-V051Super5aa	519	1859331	817	991	1111	1178
IONCAS-V051Ultra	516	1859330	816	990	1111	1178
IONCAS-V051Ultra	516	1859331	817	991	1111	1178
IONCAS-V052	55	V52gRNA	424	870	1112	1178
IONCAS-V053	56	V53gRNA	425	871	1113	1201
IONCAS-V054	57	V54gRNA	426	872	1114	1202
IONCAS-V055	58	V55gRNA	427	873	1115	1178
IONCAS-V055	58	gV55-DNM-1	623	N/A	1115	1178
IONCAS-V055	58	gV55-DNM-3	638	N/A	1115	1178
IONCAS-V055	58	gV55-FA-1	651	N/A	1115	1178
IONCAS-V055	58	gV55-FA-2	664	N/A	1115	1178
IONCAS-V055	58	gV55-NLR	678	N/A	1115	1178
IONCAS-V055	58	gV55-MS-1	691	N/A	1115	1178
IONCAS-V055	58	gV55-RUN-2	706	N/A	1115	1178
IONCAS-V055	58	gV55-RUN-3	721	N/A	1115	1178
IONCAS-V055	58	gV55-RUN-4	735	N/A	1115	1178
IONCAS-V056	59	V56gRNA	428	874	1116	1203
IONCAS-V057	60	V57gRNA	429	875	1117	1185
IONCAS-V058	61	V58gRNA	430	876	1118	1204
IONCAS-V059	62	V59gRNA	431	877	1119	1205
IONCAS-V060	63	V60gRNA	432	878	1120	1206
IONCAS-V061	64	V61gRNA	433	879	1121	1207
IONCAS-V062	65	V62gRNA	434	880	1122	1197
IONCAS-V063	66	V63gRNA	435	881	1123	1208
IONCAS-V064	67	V64gRNA	436	882	1124	1178
IONCAS-V065	68	V65gRNA	437	883	1125	1209
IONCAS-V066	69	V66gRNA	438	884	1126	1210
IONCAS-V067	70	V67gRNA	439	885	1127	1211
IONCAS-V068	71	V68gRNA	440	886	1128	1193
IONCAS-V069	72	V69gRNA	441	887	1129	1212
IONCAS-V070	73	V70gRNA	442	888	1130	1213
IONCAS-V071	74	V71gRNA	443	889	1131	1179
IONCAS-V071	74	gV71-DNM-1	624	N/A	1131	1179
IONCAS-V071	74	gV71-DNM-3	639	N/A	1131	1179
IONCAS-V071	74	gV71-FA-1	652	N/A	1131	1179
IONCAS-V071	74	gV71-FA-2	665	N/A	1131	1179
IONCAS-V071	74	gV71-NLR	679	N/A	1131	1179
IONCAS-V071	74	gV71-MS-1	692	N/A	1131	1179
IONCAS-V071	74	gV71-RUN-2	707	N/A	1131	1179
IONCAS-V071	74	gV71-RUN-3	722	N/A	1131	1179
IONCAS-V071	74	gV71-RUN-4	736	N/A	1131	1179
IONCAS-V071	74	1859339	763	937	1155	1179

IONCAS-V071	74	1859340	764	938	1155	1179
IONCAS-V071	74	1859341	765	939	1155	1179
IONCAS-V071	74	1859342	766	940	1155	1179
IONCAS-V071	74	1859343	767	941	1155	1179
IONCAS-V071	74	1859344	768	942	1155	1179
IONCAS-V071	74	1859345	769	943	1155	1179
IONCAS-V071	74	1859346	770	944	1155	1179
IONCAS-V071	74	1859660	771	945	1155	1179
IONCAS-V071	74	1859661	772	946	1155	1179
IONCAS-V071	74	1886976	1003	1037	1169	1179
IONCAS-V071	74	1886977	1004	1038	1169	1179
IONCAS-V071	74	1886978	1005	1039	1169	1179
IONCAS-V071	74	1886979	1006	1040	1169	1179
IONCAS-V071	74	1886980	1007	1041	1169	1179
IONCAS-V071	74	1886981	1008	1042	1169	1179
IONCAS-V071	74	1886982	1009	1043	1169	1179
IONCAS-V071	74	1886983	1010	1044	1169	1179
IONCAS-V071	74	1886984	1011	1045	1169	1179
IONCAS-V071	74	1886985	1012	1046	1169	1179
IONCAS-V071	74	1859328	778	952	1131	1179
IONCAS-V071	74	1859329	785	959	1131	1179
IONCAS-V071	74	1859692	779	953	1156	1179
IONCAS-V071	74	1859693	780	954	1157	1179
IONCAS-V071	74	1859704	781	955	1158	1179
IONCAS-V071	74	1859708	782	956	1159	1179
IONCAS-V071	74	1859712	783	957	1160	1179
IONCAS-V071	74	1859716	784	958	1161	1179
IONCAS-V071	74	1859694	786	960	1156	1179
IONCAS-V071	74	1859695	787	961	1157	1179
IONCAS-V071	74	1859705	788	962	1158	1179
IONCAS-V071	74	1859709	789	963	1159	1179
IONCAS-V071	74	1859713	790	964	1160	1179
IONCAS-V071	74	1859717	791	965	1161	1179
IONCAS-V071	74	1859700	808	982	1155	1179
IONCAS-V071	74	1859701	809	983	1155	1179
IONCAS-V071	74	1887453	1013	1047	1170	1228
IONCAS-V071	74	1887483	1014	1048	1171	1228
IONCAS-V071	74	1887014	1015	1049	1166	1228
IONCAS-V071	74	1887015	1016	1050	1170	1228
IONCAS-V071	74	1887016	1017	1051	1170	1228
IONCAS-V071	74	1887002	1018	1052	1171	1228
IONCAS-V071	74	1887005	1019	1053	1166	1228
IONCAS-V071	74	1887083	1020	1054	1170	1228

IONCAS-V071	74	1887113	1021	1055	1171	1228
IONCAS-V071	74	1887018	1022	1056	1166	1228
IONCAS-V071	74	1887019	1023	1057	1170	1228
IONCAS-V071	74	1887020	1024	1058	1170	1228
IONCAS-V071	74	1887008	1025	1059	1171	1228
IONCAS-V071	74	1887010	1026	1060	1166	1228
IONCAS-V071en	520	1859328	778	952	1131	1179
IONCAS-V071en	520	1859329	785	959	1131	1179
IONCAS-V071Hyper3aa	522	1859328	778	952	1131	1179
IONCAS-V071Hyper3aa	522	1859329	785	959	1131	1179
IONCAS-V071hyper4aa	524	1859328	778	952	1131	1179
IONCAS-V071hyper4aa	524	1859329	785	959	1131	1179
IONCAS-V071Plus	521	1859328	778	952	1131	1179
IONCAS-V071Plus	521	1859329	785	959	1131	1179
IONCAS-V071Super	525	1859328	778	952	1131	1179
IONCAS-V071Super	525	1859329	785	959	1131	1179
IONCAS-V071Super5aa	526	1859328	778	952	1131	1179
IONCAS-V071Super5aa	526	1859329	785	959	1131	1179
IONCAS-V071Uber	527	1859328	778	952	1131	1179
IONCAS-V071Uber	527	1859329	785	959	1131	1179
IONCAS-V071Uber7aa	528	1859328	778	952	1131	1179
IONCAS-V071Uber7aa	528	1859329	785	959	1131	1179
IONCAS-V071Ultra	523	1859328	778	952	1131	1179
IONCAS-V071Ultra	523	1859329	785	959	1131	1179
IONCAS-V072	75	V72gRNA	444	890	1132	1214
IONCAS-V073	76	V73gRNA	445	891	1133	1215
IONCAS-V073	76	V73gAS-DNM-1	625	N/A	1152	1228
IONCAS-V073	76	V73gAS-DNM-3	640	N/A	1152	1228
IONCAS-V073	76	V73gAS-FA-1	653	N/A	1152	1228
IONCAS-V073	76	V73gAS-FA-2	666	N/A	1152	1228
IONCAS-V073	76	V73gAS-NLR	680	N/A	1152	1228
IONCAS-V073	76	V73gAS-MS-1	693	N/A	1152	1228
IONCAS-V073	76	V73gAS-RUN-2	708	N/A	1152	1228
IONCAS-V073	76	V73gAS-RUN-3	723	N/A	1152	1228
IONCAS-V073	76	V73gAS-RUN-4	737	N/A	1152	1228
IONCAS-V074	77	V74gRNA	446	892	1134	1180
IONCAS-V075	78	V75gRNA	447	893	1135	1216
IONCAS-V076	79	V76gRNA	448	894	1136	1217
IONCAS-V077	80	V77gRNA	449	895	1061	1176
IONCAS-V078	81	V78gRNA	450	896	1137	1218
IONCAS-V079	82	V79gRNA	451	897	1138	1219
IONCAS-V081	83	V81gRNA	452	898	1139	1220
IONCAS-V082	84	V82gRNA	453	899	1140	1179

IONCAS-V082	84	V82gAS-DNM-1	625	N/A	1152	1228
IONCAS-V082	84	V82gAS-DNM-3	640	N/A	1152	1228
IONCAS-V082	84	V82gAS-FA-1	653	N/A	1152	1228
IONCAS-V082	84	V82gAS-FA-2	666	N/A	1152	1228
IONCAS-V082	84	V82gAS-NLR	680	N/A	1152	1228
IONCAS-V082	84	V82gAS-MS-1	693	N/A	1152	1228
IONCAS-V082	84	V82gAS-RUN-2	708	N/A	1152	1228
IONCAS-V082	84	V82gAS-RUN-3	723	N/A	1152	1228
IONCAS-V082	84	V82gAS-RUN-4	737	N/A	1152	1228
IONCAS-V082	84	1859672	738	912	1153	1228
IONCAS-V082	84	1859673	739	913	1153	1228
IONCAS-V082	84	1859674	740	914	1153	1228
IONCAS-V082	84	1859675	741	915	1153	1228
IONCAS-V082	84	1859676	742	916	1153	1228
IONCAS-V082	84	1859677	773	947	1153	1228
IONCAS-V082	84	1859678	774	948	1153	1228
IONCAS-V082	84	1859679	775	949	1153	1228
IONCAS-V082	84	1859680	776	950	1153	1228
IONCAS-V082	84	1859681	777	951	1153	1228
IONCAS-V082	84	1859332	812	986	1152	1228
IONCAS-V082	84	1859333	818	992	1152	1228
IONCAS-V082	84	1859333	818	992	1152	1228
IONCAS-V082en	529	1859332	812	986	1152	1228
IONCAS-V082en	529	1859333	818	992	1152	1228
IONCAS-V082Hyper3aa	531	1859332	812	986	1152	1228
IONCAS-V082Hyper3aa	531	1859333	818	992	1152	1228
IONCAS-V082hyper3aa	531	1859333	818	992	1152	1228
IONCAS-V082hyper4aa	533	1859332	812	986	1152	1228
IONCAS-V082hyper4aa	533	1859333	818	992	1152	1228
IONCAS-V082Plus	530	1859332	812	986	1152	1228
IONCAS-V082Plus	530	1859333	818	992	1152	1228
IONCAS-V082Ultra	532	1859332	812	986	1152	1228
IONCAS-V082Ultra	532	1859333	818	992	1152	1228
IONCAS-V083	85	V83gRNA	454	900	1141	1221
IONCAS-V084	86	V84gRNA	455	901	1142	1222
IONCAS-V085	87	V85gRNA	456	902	1143	1223
IONCAS-V085	87	V85gAS-DNM-1	625	N/A	1152	1228
IONCAS-V085	87	V85gAS-DNM-3	640	N/A	1152	1228
IONCAS-V085	87	V85gAS-FA-1	653	N/A	1152	1228
IONCAS-V085	87	V85gAS-FA-2	666	N/A	1152	1228
IONCAS-V085	87	V85gAS-NLR	680	N/A	1152	1228
IONCAS-V085	87	V85gAS-MS-1	693	N/A	1152	1228
IONCAS-V085	87	V85gAS-RUN-2	708	N/A	1152	1228

IONCAS-V085	87	V85gAS-RUN-3	723	N/A	1152	1228
IONCAS-V085	87	V85gAS-RUN-4	737	N/A	1152	1228
IONCAS-V086	88	V86gRNA	457	903	1144	1179
IONCAS-V086	88	V86gAS-DNM-1	625	N/A	1152	1228
IONCAS-V086	88	V86gAS-DNM-3	640	N/A	1152	1228
IONCAS-V086	88	V86gAS-FA-1	653	N/A	1152	1228
IONCAS-V086	88	V86gAS-FA-2	666	N/A	1152	1228
IONCAS-V086	88	V86gAS-MS-1	693	N/A	1152	1228
IONCAS-V086	88	V86gAS-RUN-2	708	N/A	1152	1228
IONCAS-V086	88	V86gAS-RUN-3	723	N/A	1152	1228
IONCAS-V086	88	V86gAS-RUN-4	737	N/A	1152	1228
IONCAS-V087	89	V87gRNA	458	904	1145	1224
IONCAS-V088	90	V88gRNA	459	905	1146	1225
IONCAS-V088	90	V88gAS-DNM-1	625	N/A	1152	1228
IONCAS-V088	90	V88gAS-DNM-3	640	N/A	1152	1228
IONCAS-V088	90	V88gAS-FA-1	653	N/A	1152	1228
IONCAS-V088	90	V88gAS-FA-2	666	N/A	1152	1228
IONCAS-V088	90	V88gAS-NLR	680	N/A	1152	1228
IONCAS-V088	90	V88gAS-MS-1	693	N/A	1152	1228
IONCAS-V088	90	V88gAS-RUN-2	708	N/A	1152	1228
IONCAS-V088	90	V88gAS-RUN-3	723	N/A	1152	1228
IONCAS-V088	90	V88gAS-RUN-4	737	N/A	1152	1228
IONCAS-V089	91	V89gRNA	460	906	1147	1226
IONCAS-V090	92	V90gRNA	461	907	1148	1227
IONCAS-V091	93	V91gRNA	462	908	1149	1178
IONCAS-V092	94	V92gRNA	463	909	1150	1178
IONCAS-V093	95	V93gRNA	464	910	1151	1180

CLAIMS

1. A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids, wherein the amino sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length portion of any of SEQ ID NOs: 4-95, 501-533, 600-601.
2. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 4-95, 501-533, 600-601.
3. A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids, wherein the amino sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length portion of any of SEQ ID NOs: 18, 20-24, 26-32, 34, 36-39, 41, 42, 44, 45, 50, 53-58, 75-76, 84, 85, 87-91, or 95.
4. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 18, 20-24, 26-32, 34, 36-39, 41, 42, 44, 45, 50, 53-58, 75-76, 84, 85, 87-91, or 95.
5. The guided nucleic acid binding agent of any of claims 1-4, consisting of the first polypeptide.
6. The guided nucleic acid binding agent of any of claims 1-5, wherein the first polypeptide is a Cas protein.
7. The guided nucleic acid binding agent of any of claims 1-6, wherein the first polypeptide comprises a RuvC domain.
8. The guided nucleic acid binding agent of any of claims 1-7, wherein the first polypeptide has exactly one catalytically active nuclease site.
9. The guided nucleic acid binding agent of claim 8, wherein the first polypeptide can introduce a double-stranded break in DNA.
10. The guided nucleic acid binding agent of any of claims 1-7, wherein the first polypeptide has zero catalytically active nuclease sites.
11. The guided nucleic acid binding polypeptide of claim 10, wherein first polypeptide comprises a RuvC domain and the RuvC domain contains an inactivating substitution of a conserved Asp.
12. The guided nucleic acid binding agent of any preceding claim, comprising a second polypeptide.

13. The guided nucleic acid binding agent of claim 12, wherein the second polypeptide comprises a heterologous domain.
14. The guided nucleic acid binding agent of claim 13, wherein the functional domain is selected from a transcriptional activator, a transcriptional repressor, a methyltransferase, a demethylase, a deaminase, an acetyltransferase, or a deacetylase.
15. The guided nucleic acid binding agent of any of claims 12-14, wherein the first polypeptide and the second polypeptide are fused to form a single protein.
16. The guided nucleic acid binding agent of claim 15, wherein the second polypeptide is fused to the N-terminus of the first polypeptide, forming a fusion protein.
17. The guided nucleic acid binding agent of claim 15, wherein the second polypeptide is fused to the C-terminus of the first polypeptide, forming a fusion protein.
18. The guided nucleic acid binding agent of any of any of claims 12-14, further comprising one or more additional fused heterologous domains.
19. The guided nucleic acid binding agent of claims 15-18, wherein the fusion protein is an epigenetic editing protein.
20. A guide, wherein the guide comprises a protein-recognition region, wherein the protein-recognition region binds to the first polypeptide of any of claims 1-19.
21. The guide of claim 20, wherein the guide is a single guide comprising an oligonucleotide, and the protein-recognition region comprises a direct repeat region.
22. The guide of claim 21, wherein the nucleobase sequence of the direct repeat region of the guide has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 1176-1228.
23. The guide any of claims 20-22, wherein the guide comprises a modified oligonucleotide.
24. The guide of claim 23, wherein the guide comprises at least one modified sugar moiety.
25. The guide of claim 24, wherein the modified sugar moiety is selected from a 2'-OMe and a 2'-F.
26. The guide of claim 25, wherein the guide comprises a 2'-OMe or 2'-F sugar moiety within the first five nucleosides of the 5' end or the last five nucleosides 3' end of the guide.
27. The guide of any of claims 23-26, wherein the guide comprises a modified internucleoside linkage.
28. The guide of claim 27, wherein the modified internucleoside linkage is a phosphorothioate internucleoside linkage.
29. The guide of claim 27 or 28, wherein the guide comprises at least one modified internucleoside linkage within the first five nucleosides of the 5' end or the last five nucleosides at the 3' end of the guide.
30. The guide of any of claims 20-29, wherein the guide consists of an oligonucleotide.
31. The guide of any of claims 20-30, wherein the guide comprises a target-recognition region that is at least 90%, at least 95%, or 100% complementary to a target sequence.

32. The guide of any of claims 20-31, wherein the guide comprises a stabilizing region.
33. The guide of claim 32, wherein the stabilizing region is at the 5'-end of the guide.
34. The guide of claim 32, wherein the stabilizing region is at the 3'-end of the guide.
35. The guide of any of claims 32-34, wherein the stabilizing region has a sequence that is 90%, 95%, or 100% identical to any of SEQ ID Nos: 488-494, 498, or 1172-1175.
36. An editing system comprising the nucleic acid binding agent of any of claims 1-19 and a guide of any of claims 20-35.
37. An editing system comprising:
    a. A guided nucleic acid binding agent, comprising at least a first polypeptide, wherein the amino acid sequence of the polypeptide has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the full length of any of SEQ ID NOs: 4-95, 501-533, or 600-601; and
    b. A guide, wherein the nucleobase sequence of the direct repeat region of the guide has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to SEQ ID NO: 1176-1228;

    wherein the SEQ ID NO: of the first polypeptide and the SEQ ID NO: of the guide are selected from the same row of Table 52.
38. The editing system of claim 37, wherein the guided nucleic acid binding agent further comprises a second polypeptide that comprises at least one heterologous domain.
39. The editing system of claim 37 or 38, comprising a ribonucleoprotein (RNP), wherein the RNP comprises the guided nucleic acid binding agent and the guide.
40. A nucleic acid encoding the first polypeptide of any of claims 1-19.
41. An expression system comprising:
    a. a nucleic acid of claim 40; and
    b. a guide of any of claims 20-35.
42. An expression system comprising:
    a. a nucleic acid of claim 40; and
    b. a nucleic acid encoding the guide of any of claims 20-22.
43. The nucleic acid of claim 40 or expression system of claim 41 or 42, wherein the nucleic acid encoding the first polypeptide is an exogenous mRNA.
44. The nucleic acid of claim 40 or expression system of claim 41 or 42, wherein the nucleic acid encoding the first polypeptide is a DNA.
45. The expression system of any of claims 41-44, wherein the nucleic acid encoding the guide is an unmodified RNA.
46. The expression system of any of claims 41-44, wherein the nucleic acid encoding the guide is a DNA.

47. The expression system of claim 42, wherein the polypeptide and the guide are encoded by the same nucleic acid.
48. An LNP at least partially encapsulating the editing system of any of claims 37-39, the nucleic acid of any of claims 40, 43 or 44, or the expression system of any of claims 41, 42, or 45-47.
49. A viral vector comprising the nucleic acid of any of claims 40, 43 or 44, or the expression system of any of claims 41, 42, or 45-47.
50. A composition comprising:
    a. the guided nucleic acid binding agent of any of claims 1-19 and the guide of any of claims 20-35;
    b. the editing system of any of claims 36-39;
    c. the expression system of any of claims 41, 42, or 45-47;
    d. the LNP of claim 48; or
    e. the viral vector of claim 49.
51. A method of editing a target nucleic acid, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
52. A method of creating a discontinuity in a target nucleic acid, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
53. A method of creating a double stranded break or a nick in a target DNA, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
54. A method of gene silencing, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
55. A method of gene activation, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
56. A method of homologous gene repair, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
57. A method of gene visualization, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
58. A method of epigenetic modulation of gene expression, comprising administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 to a subject.
59. A method of editing a target nucleic acid, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
60. A method of creating a discontinuity in a target nucleic acid, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
61. A method of creating a double stranded break or a nick in a target DNA, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.

62. A method of gene silencing, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
63. A method of gene activation, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
64. A method of homologous gene repair, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
65. A method of gene visualization, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
66. A method of epigenetic modulation of gene expression, comprising contacting a cell with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
67. The method of any of claims 51, 52, 59 or 60, wherein the target nucleic acid is a target DNA.
68. The method of any of claims 53, 61 or 67, wherein the non-complementary strand of the target DNA comprises a sequence selected from Table 6 within 18-20 nucleobases from the cut site in the non-complementary strand.
69. The method of any of claims 59-66, wherein the cell is in a subject.
70. The LNP of claim 48, viral vector of claim 49, or composition of claim 50 for use in therapy.
71. A complex comprising the editing system of claim 37 in contact with a target DNA, wherein the complementary strand of the DNA comprises a sequence complementary to the guide adjacent to a sequence complementary to a sequence selected from Table 6.
72. A method of treating a disease or disorder in a subject, wherein the method comprises administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
73. A method of autologous cell therapy, wherein the method comprises providing an autologous cell, contacting an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 with the cell, and administering the cell to a subject.
74. A method of allogeneic cell therapy, wherein the method comprises providing an allogeneic cell, contacting an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 with the cell, and administering the cell to a subject.
75. A cell comprising an edited target nucleic acid, produced by the method of claim 59.
76. A cell comprising a discontinuity in a target nucleic acid, produced by the method of claim 60.
77. A cell comprising a double stranded break or a nick in a target DNA, produced by the method of claim 61.
78. A cell comprising a silenced gene, produced by the method of claim 62.
79. A cell comprising an activated gene, produced by the method of claim 63.
80. A cell comprising a homologously repaired gene, produced by the method of claim 64.
81. A cell comprising a gene that can be visualized, produced by the method of claim 65.

82. A gene, the expression of which has been epigenetically modified, produced by the method of claim 66.
83. Use of an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 in the manufacture of a medicament for treating a disease or disorder in a subject.
84. Use of a cell of any of claims 75-82 in the manufacture of a medicament for treating a disease or disorder in a subject.
85. Use of an autologous cell, wherein the cell has been contacted with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 in the manufacture of a medicament for autologous cell therapy.
86. Use of an allogeneic cell, wherein the cell has been contacted with an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50 in the manufacture of a medicament for allogeneic cell therapy.
87. The cell of any of claims 75-81 for use in therapy.
88. A method of diagnosing a disease or disorder in a subject, wherein the method comprises administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
89. A method of assessing responsiveness to a treatment of a disease or disorder in a subject, wherein the method comprises administering an LNP of claim 48, a viral vector of claim 49, or a composition of claim 50.
90. A kit comprising an LNP of claim 48, a viral vector of claim 49, a composition of claim 50, or a cell of any of claims 75-82 and optionally instructions for use, optionally wherein the kit is for, or when used for, treating a disease or disorder in a subject or a cell thereof.
91. A compound comprising an oligonucleotide consisting of 25 to 150 linked nucleosides, wherein the 5'-terminus of the oligonucleotide consists of 15 or 16 nucleotides having a sequence with at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length region of SEQ ID NOs: 1172-1174.
92. The compound of claim 91, wherein the oligonucleotide comprises at least one modified nucleoside or at least one modified internucleoside linkage.
93. The compound of claim 91 or 92, wherein the compound comprises a direct repeat region.
94. The compound of any of claims 91-93, wherein the compound comprises a target-recognition region.
95. The compound of any of claims 91-94, wherein the target-recognition region is complementary to a mammalian target.
96. The compound of any of claims 91-95, wherein the compound is a guide, and wherein the gene editing activity of the guide is increased relative to the gene editing activity of an otherwise identical guide that is truncated by 15 or 16 nucleotides at the 5'-end compared to the compound of any of embodiments 147-151.
97. An editing system comprising the compound of any of claims 91-96.

FIG. 1

In vitro cleavage activity of ION-CAS proteins in HEK293T nuclear extract using their corresponding Cas-specific guide RNA

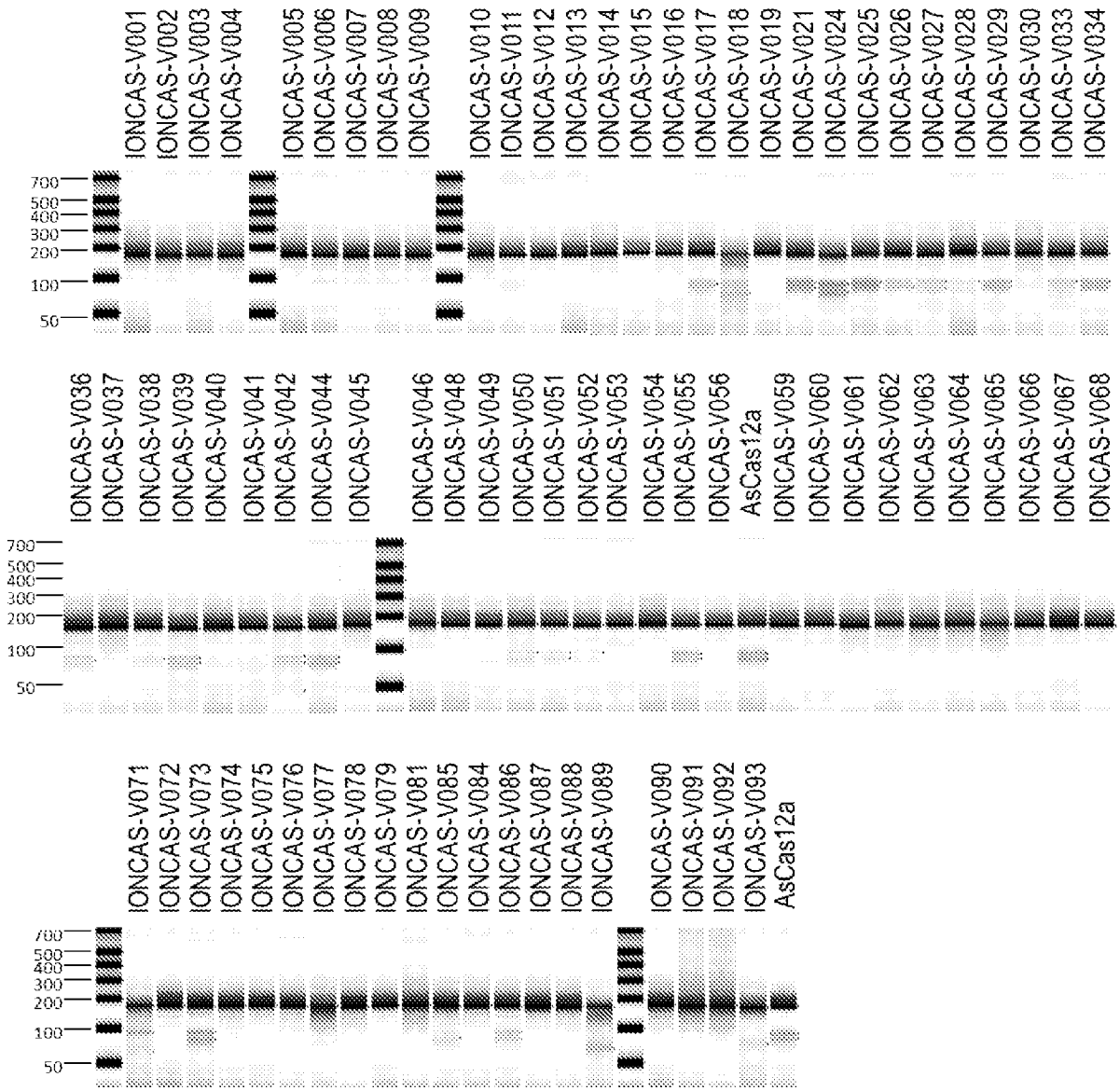


FIG. 2A

In vitro cleavage activity of ION-CAS proteins in HEK293T nuclear extracts using an AsCas12a guide RNA.

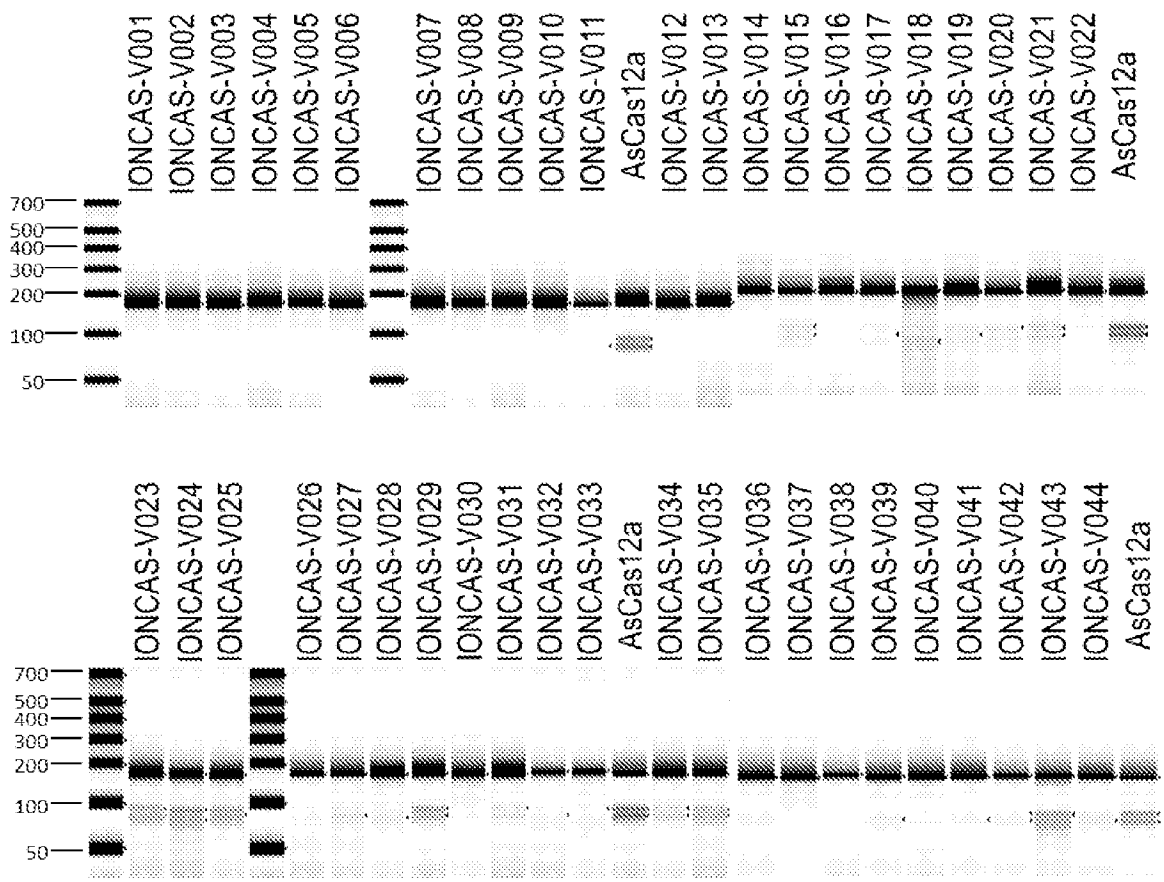


FIG. 2B

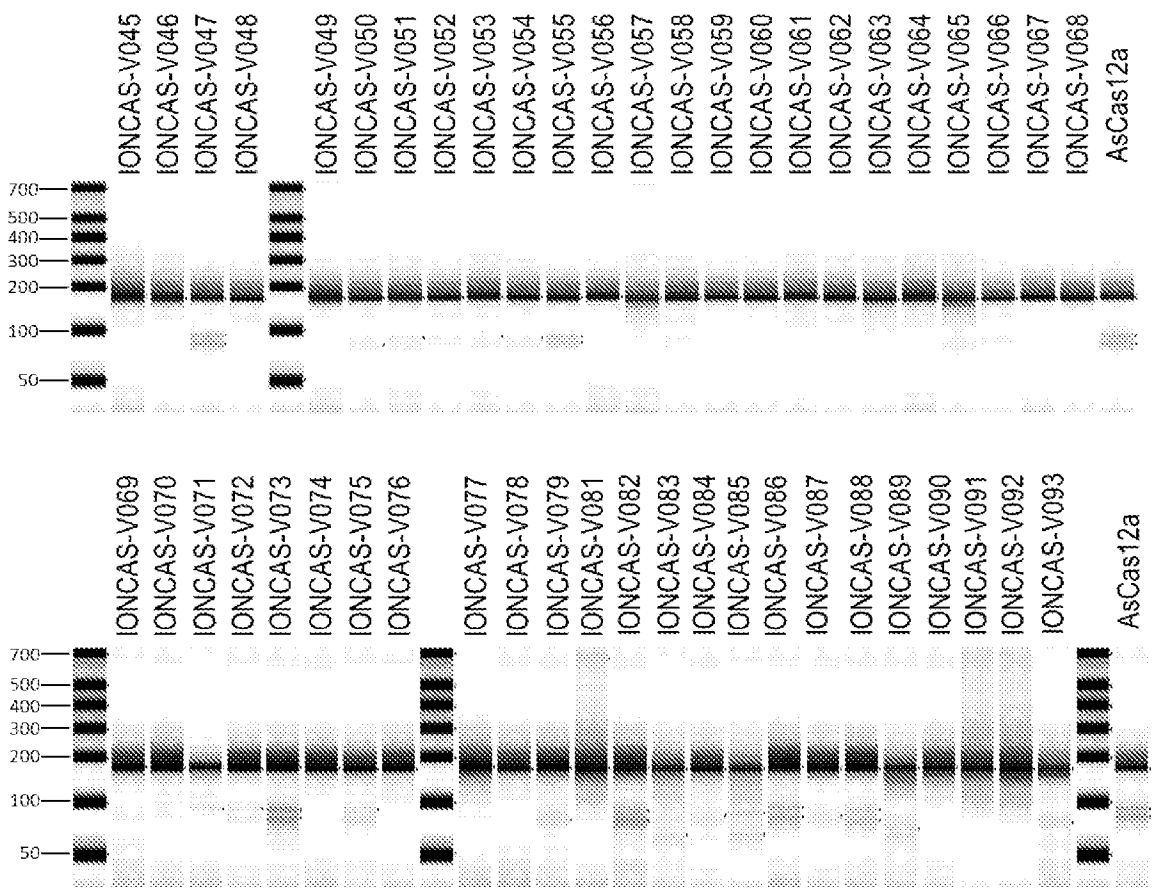
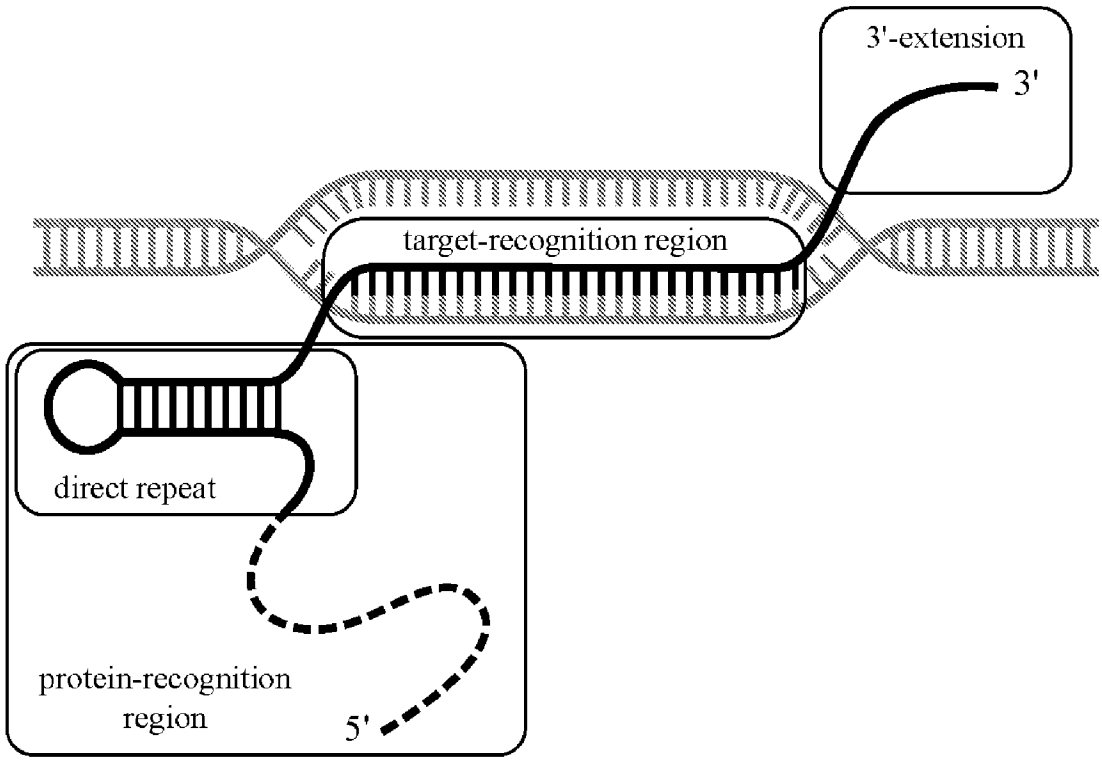

IONCAS-V045
IONCAS-V046
IONCAS-V047
IONCAS-V048
IONCAS-V049
IONCAS-V050
IONCAS-V051
IONCAS-V052
IONCAS-V053
IONCAS-V054
IONCAS-V055
IONCAS-V056
IONCAS-V057
IONCAS-V058
IONCAS-V059
IONCAS-V060
IONCAS-V061
IONCAS-V062
IONCAS-V063
IONCAS-V064
IONCAS-V065
IONCAS-V066
IONCAS-V067
IONCAS-V068
AsCas12a
700
500
400
300
200
100
50
IONCAS-V069
IONCAS-V070
IONCAS-V071
IONCAS-V072
IONCAS-V073
IONCAS-V074
IONCAS-V075
IONCAS-V076
IONCAS-V077
IONCAS-V078
IONCAS-V079
IONCAS-V081
IONCAS-V082
IONCAS-V083
IONCAS-V084
IONCAS-V085
IONCAS-V086
IONCAS-V087
IONCAS-V088
IONCAS-V089
IONCAS-V090
IONCAS-V091
IONCAS-V092
IONCAS-V093
AsCas12a
700
500
400
300
200
100
50

FIG 3.


3'-extension
3'
target-recognition region
direct repeat
protein-recognition region
5'

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/061295

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed.
   b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),
      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/061295

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

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

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

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

3. ☑ Claims Nos.: **6-36, 40-70, 72-97**
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-5, 37-39, and 71 are drawn to guided nucleic acid binding agents, editing systems, and complexes.

The first invention of Group I+ is restricted to a guided nucleic acid binding agent comprising SEQ ID NO: 4, a guide comprising SEQ ID NO: 373, editing systems comprising the same, and complexes comprising the same. The first named invention has been selected based on the guidance set forth in section 10.54 of the PCT International Search and Preliminary Examination Guidelines. Specifically, the first named invention was selected based on the first listed compound species presented in the claims (see claim 1). It is believed that claims 1, 2, 5, 37-39, and 71 read on this first named invention and thus these claims will be searched without fee to the extent that they read on a guided nucleic acid binding agent comprising SEQ ID NO: 4, a guide comprising SEQ ID NO: 373, editing systems comprising the same, and complexes comprising the same.

Applicant is invited to elect additional guided nucleic acid binding agents and guides to be searched in a specific combination by paying an additional fee for each set of elections. An exemplary election would be a guided nucleic acid binding agent comprising SEQ ID NO: 1, a guide comprising SEQ ID NO: 465, editing systems comprising the same, and complexes comprising the same. Additional guided nucleic acid binding agents and guides will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The Groups I+ formulas do not share a significant structural element responsible for forming editing systems and complexes requiring the selection of alternative guided nucleic acid binding agents and guides where "a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids, wherein the amino sequence of the polypeptide has at least 85%, at least 90%, at

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

least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to an equal length portion of any of SEQ ID NOs: 4-95, 501-533, 600-601."

Additionally, even if Groups I+ were considered to share the technical features of a guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids; A guided nucleic acid binding agent, comprising at least a first polypeptide; A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids; A guided nucleic acid binding agent, comprising at least a first polypeptide; An editing system comprising: a. A guided nucleic acid binding agent, comprising at least a first polypeptide; and b. A guide; and A complex comprising the editing system in contact with a target DNA, wherein the complementary strand of the DNA comprises a sequence complementary to the guide adjacent to a sequence complementary to a sequence, these shared technical features do not represent a contribution over the prior art.

Specifically, WO 2023/056291 A1 to Acrigen Biosciences (hereinafter, "Acrigen") teaches a guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids (a nuclease comprising an amino acid sequence having at least 70% identity to any of SEQ ID NOs: 1-1096 and at least one gRNA, Abstract; SEQ ID NO: 235 of Acrigen consists of at least 1300 linked amino acids); A guided nucleic acid binding agent, comprising at least a first polypeptide (a nuclease comprising an amino acid sequence having at least 70% identity to any of SEQ ID NOs: 1-1096 and at least one gRNA, Abstract; SEQ ID NO: 235 of Acrigen consists of at least 1300 linked amino acids); A guided nucleic acid binding agent, comprising at least a first polypeptide consisting of at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, or at least 1300 linked amino acids (a nuclease comprising an amino acid sequence having at least 70% identity to any of SEQ ID NOs: 1-1096 and at least one gRNA, Abstract; SEQ ID NO: 235 of Acrigen consists of at least 1300 linked amino acids); A guided nucleic acid binding agent, comprising at least a first polypeptide (a nuclease comprising an amino acid sequence having at least 70% identity to any of SEQ ID NOs: 1-1096 and at least one gRNA, Abstract; SEQ ID NO: 235 of Acrigen consists of at least 1300 linked amino acids); An editing system comprising: a. A guided nucleic acid binding agent, comprising at least a first polypeptide; and b. A guide (system comprising a nuclease comprising an amino acid sequence having at least 70% identity to any of SEQ ID NOs: 1-1096 and at least one gRNA, Abstract; SEQ ID NO: 235 of Acrigen consists of at least 1300 linked amino acids); and A complex comprising the editing system in contact with a target DNA, wherein the complementary strand of the DNA comprises a sequence complementary to the guide adjacent to a sequence complementary to a sequence (the nuclease may be delivered as an RNP (ribonucleoprotein complex) in which it is already complexed with an appropriate guide RNA, Para. [0272]; The target sequence may or may not be flanked by a protospacer adjacent motif (PAM) sequence. In certain embodiments, a nucleic acid-guided nuclease can only cleave a target sequence if an appropriate PAM is present, Para. [0234]).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/061295

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☑ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1, 2, 5, 37-39, 71**

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/061295

A. CLASSIFICATION OF SUBJECT MATTER

IPC: ***C12N 9/22*** (2025.01); ***A61K 38/46*** (2025.01); ***A61K 9/127*** (2025.01); ***C07K 14/195*** (2025.01); ***C12N 15/10*** (2025.01); ***C12N 15/113*** (2025.01); ***C12N 15/52*** (2025.01); ***C12N 15/86*** (2025.01); *C12N 5/10* (2025.01); *A61K 35/12* (2025.01)

CPC: ***C12N 9/22; C12N 15/102; C12N 15/113; C12N 15/52; C12N 15/86; A61K 38/465; A61K 9/127; C07K 14/195;*** *C12N 2310/20; A61K 35/12; C07K 2319/80; C12N 5/10*

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

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

See Search History Document

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

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2023/056291 A1 (ACRIGEN BIOSCIENCES) 06 April 2023 (06.04.2023) entire document	1, 2, 5, 37-39, 71
A	US 2023/0235362 A1 (ARTISAN DEVELOPMENT LABS INC.) 27 July 2023 (27.07.2023) entire document	1, 2, 5, 37-39, 71
A	US 2023/0357796 A1 (DANMARKS TEKNISKE UNIVERSITET) 09 November 2023 (09.11.2023) entire document	1, 2, 5, 37-39, 71

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
13 February 2025 (13.02.2025)	**24 April 2025 (24.04.2025)**

Name and mailing address of the ISA/US	Authorized officer
COMMISSIONER FOR PATENTS **MAIL STOP PCT, ATTN: ISA/US** **P.O. Box 1450** **Alexandria, VA 22313-1450** **UNITED STATES OF AMERICA**	**TAINA MATOS**
Facsimile No. **571-273-8300**	Telephone No. **571-272-4300**