DELIVERY SYSTEM FOR FUNCTIONAL NUCLEASES

RELATED APPLICATION

[0001] This application claims priority under 35 U.S.C. § 365(c) to U.S. application, U.S.S.N. 14/462,189, filed August 18, 2014, and to U.S. application, U.S.S.N. 14/462,163, filed August 18, 2014, and also claims priority under 35 U.S.C. § 119(e) to U.S. provisional patent application, U.S.S.N. 61/874,746, filed September 6, 2013, each of which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Macromolecular delivery into mammalian cells is an attractive approach for cell manipulation, as it would allow modulation of gene expression and modification of the genome, which, in turn, would open new avenues for research and enable the therapeutic targeting of molecules currently viewed as "undruggable" by small molecules. In particular, recombinant nucleases targeting genes or alleles associated with disease have great potential as therapeutic agents. The current methods of macromolecular delivery include viral delivery of nucleic acid molecules, receptor-mediated delivery of nucleic acids or proteins, and the use of protein fusions with cell-penetrating peptides such as TAT, Arg9, or Penetratin for the delivery of proteins. Each of these delivery systems offers benefits for particular applications; in most cases, however, questions regarding efficacy, cytotoxicity, and ease of preparation remain. Easily prepared reagents capable of effectively delivering macromolecules (*e.g.*, functional effector proteins) to a variety of cell lines without significant cytotoxicity or other adverse side effect remain of considerable concern.

[0003] Most proteins do not spontaneously enter mammalian cells and are thus naturally limited in their use as research tools and their potential as therapeutic agents. Techniques for the delivery of proteins into mammalian cells have been developed recently to address intracellular targets. These techniques include the use of lipid-based reagents (Zelphati *et al.*, *J. Biol. Chem.* 276, 35103-35110, 2001), nanoparticles (Hasadsri *et al.*, *J. Biol. Chem.*, 2009), vault ribonucleoprotein particles (Lai *et al.*, *ACS Nano* 3, 691-699, 2009); genetic or chemical fusion to receptor ligands (Gabel *et al.*, *J. Cell Biol.* 103, 1817-1827, 1986; Rizk *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 106, 11011-11015, 2009); and fusion to cell-penetrating peptides (Wadia *et al.*, Curr. Protein Pept. Sci. 4, 97-104, 2003; Zhou *et al.*, Cell Stem Cell 4, 381-384, 2009). Perhaps the most common method for protein delivery is genetic fusion to protein transduction domains (PTDs) including the HIV-1 transactivator of

transcription (Tat) peptide and polyarginine peptides. These cationic PTDs promote association with negatively charged cell-surface structures and subsequent endocytosis of exogenous proteins. Both Tat and polyarginine have been used to deliver a variety of macromolecules into cells both *in vitro* and *in vivo* (Wadia *et al.*, *Curr. Protein Pept. Sci.* 4, 97-104, 2003; Zhou *et al.*, *Cell Stem Cell* 4, 381-384, 2009; Myou *et al.*, *J. Immunol.* 169, 2670-2676, 2002; Bae *et al.*, *Clin. Exp. Immunol.* 157, 128-138, 2009; Schwarze *et al.*, *Science* 285, 1569-1572, 1999). Despite these advances, intracellular targets remain difficult to affect using exogenous proteins, and even modest success can require toxic concentrations of the respective transduction agent due to the low efficiency with which proteins are functionally delivered into cells (Zhou *et al.*, *Cell Stem Cell* 4, 381-384, 2009; Wang *et al.*, *Nat. Biotechnol.* 26, 901-908, 2008). Therefore, there remains a need for better delivery systems for getting functional effector proteins into cells to target intracellular biomolecules.

SUMMARY OF THE INVENTION

[0004] The present disclosure provides novel systems, compositions, preparations, kits, and related methods for delivering functional effector proteins, such as, for example, site-specific proteins that bind nucleic acids, into cells using a supercharged protein (*e.g.*, a positively charged supercharged protein), a cationic polymer, or a cationic lipid. In some embodiments, the nucleases are TALE nucleases, RNA-programmable nucleases or engineered RNA-programmable genome-editing enzymes (such as Cas9 and variants or fusions thereof), or zinc finger nucleases. In some embodiments, the transcription factors are TALE transcriptional activators or repressors. In some embodiments, the effector proteins are recombinases. As described in greater detail herein, fusing or associating functional effector proteins (*e.g.*, nucleases, transcriptional activators/repressors, Cas9 proteins including variants and fusions thereof, *etc.*) with positively charged supercharged proteins allows for delivery of the proteins to the interior of cells, for example to affect gene expression or genomic modifications. It was also found that fusing or associating functional effector proteins with negatively charged supercharged proteins allows for the proteins to assocoaite with cationic lipids or cationic polymers, which provides potent delivery of the proteins to the interior of a cell. Further, functional effector proteins that are naturally negatively charged (*e.g.*, VP64 transcriptional activators, the anionic 3xFLAG peptide tag, and fusions thereof) or functional effector proteins (*e.g.*, Cas9 proteins, and variants and fusions thereof) that associate with nucleic acids (*e.g.*, guide RNAs; "gRNAs") which are

inherently negatively charged, can associate with cationic lipids or cationic polymers for delivery to cells (*e.g.*, in the absence of a supercharged protein).

[0005] While delivery of effector proteins has proven effective for extracellular targets, their use to address intracellular targets is comparatively undeveloped due to the inability of most proteins to spontaneously enter mammalian cells. Enabling exogenous proteins to access intracellular targets is most commonly achieved by delivery of their encoding DNA sequences through chemical transfection, electroporation, or viral delivery. The introduction of exogenous DNA into cells, however, raises the possibility of permanent recombination into the genome, potential disruption of endogenous genes, and long-term exposure to the encoded agent. For some research or therapeutic applications, including genome editing applications that seek to effect a one-time, permanent modification of genomic DNA, the functional delivery of non-replicable protein agents may offer improved safety or broader applicability. Further, while the delivery of proteins using cationic compounds such as lipids and polymers has remained technically challenging and in many cases induces cellular toxicity, it was surprisingly found, using the compositions and methods provided herein, that certain functional effector proteins (*e.g.,* Cas9 proteins and variants and fusions thereof, recombinases, transcriptional activators/repressors, *etc.*) can be delivered to cells with no or minimal toxicity, in some cases mediating genomic modifications with significant improvements in efficiency and reduced off-target effects. For example, as described in Example 7, delivery of Cas9:gRNA complexes with cationic lipids is highly efficient (up to 80% modification of cultured human cells from a single treatment) and also induces higher genome modification specificity compared with plasmid transfection, typically resulting in >10-fold higher on-target:off-target DNA modification ratios in human cells.

[0006] Accordingly, in one aspect, supercharged proteins are used to deliver functional effector proteins into cells, for example nucleases, transcriptional activators/repressors, Cas9 proteins (including fusions and variants thereof), *etc*. In some embodiments, the supercharged protein has been engineered to exhibit an increase in its overall surface charge as compared to the corresponding unmodified protein. In other embodiments, the supercharged protein has been engineered to exhibit a decrease in its overall surface charge as compared to the corresponding unmodified protein. In other embodiments, the supercharged protein used in the context of this disclosure is a naturally occurring supercharged protein. The supercharged protein may be associated with the protein to be delivered through covalent or non-covalent interactions. Without wishing to be bound

by any particular theory, the Cas9 protein, variant, or fusion protein associated with a gRNA has net negative charged facilitating association with a positively charged supercharged protein. In certain embodiments, the functional effector protein associated with the supercharged protein is further associated with a cationic polymer or cationic lipid to form a composition suitable for delivery into a cell. Examples of suitable engineered or naturally occurring supercharged proteins are described in international PCT patent application PCT/US07/70254, filed June 1, 2007, published as WO 2007/143574 on December 13, 2007; in international PCT application PCT/US09/041984, filed on April 28, 2009, published as WO 2009/134808 on November 5, 2009; and in international PCT application PCT/US10/001250, filed on April 28, 2010, published as WO 2010/129023 on November 11, 2010; the entire contents of each of which are incorporated herein by reference. Further examples of supercharged proteins for use in delivering nucleases to cells are described herein. Additional examples of suitable functional effector proteins, for example, nucleases and RNA-programmable effector proteins such as Cas9 proteins, are described in U.S. Provisional Patent Application, U.S.S.N. 61/868,846, filed August 22, 2013, entitled "Engineered Transcription Activator-Like Effector (TALE) Domains and Uses Thereof," U.S. Provisional Patent Application, U.S.S.N. 61/874,609, filed September 6, 2013, entitled "Cas9 Variants and Uses Thereof," U.S. Provisional Patent Application, U.S.S.N. 61/874,682, filed September 6, 2013, entitled "Switchable Cas9 Nucleases and Uses Thereof," U.S. Non-provisional Application, U.S.S.N. 14/320,519, filed June 20, 2014, entitled "Engineered Transcription Activator-Like Effector (TALE) Domains and Uses Thereof," U.S. Non-provisional Application, U.S.S.N. 14/320,498, filed June 30, 2014, entitled "Cas9-FokI Fusion Proteins And Uses Thereof," U.S. Non-provisional Application, U.S.S.N. 14/320,467, filed june 30, 2014, entitled "Cas9-Recombinase Fusion Proteins And Uses Thereof," U.S. Non-provisional Application, U.S.S.N. 14/326,329, filed July 8, 2014, entitled "Switchable gRNAs Comprising Aptamers," U.S. Non-provisional Application, U.S.S.N. 14/326,340, filed July 8, 2014, entitled "mRNA-Sensing Switchable gRNAs," U.S. Non-provisional Application, U.S.S.N. 14/326,361, filed July 8, 2014, entitled "Extended DNA-Sensing gRNAs," U.S. Non-provisional Application, U.S.S.N. 14/325,815, filed July 8, 2014, entitled "Fusions Of Cas9 Domains And Nucleic Acid-Editing Domains," U.S. Non-provisional Application, U.S.S.N. 14/326,109, filed July 8, 2014, entitled "Methods For Nucleic Acid Editing," U.S. Non-provisional Application, U.S.S.N. 14/326,140, filed July 8, 2014, entitled "Methods For Correcting PI3K Point Mutations," U.S. Non-provisional Application, U.S.S.N. 14/326,269, filed July 9, 2014, entitled "Methods For Correcting

Presenilin Point Mutations," U.S. Non-provisional Application, U.S.S.N. 14/326,290, filed July 8, 2014, entitled "Methods For Correcting α-Antitrypsin Point Mutations," U.S. Non-provisional Application, U.S.S.N. 14/326,318, filed July 8, 2014, entitled "Methods For Correcting Von Willebrand Factor Point Mutations," U.S. Non-provisional Application, U.S.S.N. 14/326,303, filed july 8, 2014, entitled "Methods For Correcting Caspase-9 Point Mutations," and U.S. Provisional Application, U.S.S.N. 62/030,943, entitled "Cas9 Proteins Including Ligand-Dependent Inteins," the entire contents of each of which are incorporated herein by reference.

[0007] In some embodiments, the supercharged protein, engineered or naturally occurring, is positively charged. In other embodiments, for example those involving delivery of certain effector proteins using cationic lipids and/or cationic polymers, the supercharged protein is negatively charged. In certain embodiments, superpositively or supernegatively charged proteins is non-covalently associated with an effector protein. Alternatively, superpositively or supernegatively charged proteins may be covalently bound to the effector protein. In some embodiments, the effector protein is fused to a supercharged protein. In certain embodiments, the resulting fusion protein comprises a linker, *e.g.*, a cleavable linker, between the supercharged protein and the effector protein.

[0008] Some aspects of this disclosure provide compositions comprising a supercharged protein associated with a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*). In some embodiments, the composition further comprises a cationic lipid. In some embodiments, the composition further comprises a cationic polymer. In some embodiments, the composition further comprises a buffer or excipient. In some embodiments, the supercharged protein has an overall positive charge that is greater than its corresponding unmodified protein and is in a quantity sufficient for and is formulated for delivery to and penetration into a cell. In other embodiments, for example those involving delivery of certain effector proteins using cationic lipids and/or cationic polymers, the supercharged protein has an overall negative charge that is greater than its corresponding unmodified protein. In some embodiments, the functional effector protein is a site-specific enzyme, *e.g.*, a nuclease, Cas9 protein, recombinase, *etc.* In some embodiments, the Cas9 protein is a wild type Cas9 protein, a Cas9 nickase, or comprises a nuclease inactivated (dCas9) protein. In some embodiments, the Cas9 protein is a fusion protein comprising dCas9. In some embodiments, the fusion protein comprises a transcriptional activator (*e.g.*, VP64), a transcriptional repressor (*e.g.*, KRAB, SID) a nuclease domain (*e.g.*, FokI), a

recombinase domain (*e.g.*, Hin, Gin, or Tn3), a deaminase (*e.g.*, a cytidine deaminase or an adenosine deaminase) or an epigenetic modifier domain (*e.g.*, TET1). In some embodiments involving nucleases, the nuclease is a TALE nuclease, a Cas9 nuclease, a Cas9 nickase, or a zinc finger nuclease. In some embodiments, the nuclease specifically binds and cleaves a nucleic acid sequence. In some embodiments, the targeted nucleic acid sequence is a sequence of a gene that is a therapeutic target, for example a gene that is desirable to inactivate in the treatment of a disease. In some embodiments, the targeted nucleic acid sequence is a PRDM16, PPARγ, VEGF-A, Oct-4, PI3K, presenilin, α-antitrypsin, von willebrand factor, or caspase-9 gene sequence.

[0009] In some embodiments, the functional effector protein is a transcription factor. In some embodiments, the functional effector protein is a TALE transcriptional activator or repressor. In some embodiments, the transcription factor, transcriptional activator, or transcriptional repressor specifically binds and activates or represses a gene. In some embodiments, the gene is a therapeutic target. In some embodiments, the functional effector protein is a TALE effector. In some embodiments, the supercharged protein is covalently bound to the functional effector protein, thus forming a fusion protein. In some embodiments, the supercharged protein is associated with the functional effector protein via a linker. In some embodiments, the linker is a cleavable linker. In some embodiments, the linker is a UV-cleavable linker or a linker that is cleaved by a lysosomal enzyme. In some embodiments, the supercharged protein is non-covalently associated with the functional effector protein, thus forming a complex. In some embodiments, the supercharged protein has an overall net positive charge. In other embodiments the supercharged protein has an overall net negative charge, and the protein(s) are associated with a cationic lipid. In other embodiments the supercharged protein has an overall net negative charge, and the protein(s) are associated with a cationic polymer. In some embodiments, the overall net positive charge is between about +5 and about +40, or the overall net negative charge is between about -5 and about -50. In some embodiments, the supercharged protein is more positively charged or is more negatively charged at physiological pH than its corresponding unmodified protein. In some embodiments, the corresponding unmodified protein is a naturally occurring protein. In some embodiments, the supercharged protein is at least +5 more positively or is at least -5 more negatively charged at physiological pH than its corresponding unmodified protein. In some embodiments, the supercharged protein is a fluorescent protein. In some embodiments, the supercharged protein is green fluorescent protein (GFP). In some embodiments, the supercharged protein is a superpositively charged GFP. In some embodiments, the

supercharged protein is a superpositively charged GFP (+36 GFP) comprising at least 20 contiguous amino acid residues of the sequence:
GGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAE
VKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAKFKIR
HNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMVLLEFV
TAAGIKHGRDERYK (SEQ ID NO: 1).
In some embodiments, the supercharged protein comprises the amino acid sequence set forth in SEQ ID NO: 1. In some embodiments, the supercharged protein consists of the amino acid sequence set forth in SEQ ID NO: 1. In some embodiments, the composition is a pharmaceutical composition. In some embodiments, the composition comprises a pharmaceutically acceptable excipient. In some embodiments, the composition is formulated for administration to a subject and comprises the supercharged protein and the functional effector protein in an amount effective for delivery to at least one cell of the subject. In some embodiments, the composition comprises the supercharged protein and the functional effector protein in an amount effective for inducing a measurable therapeutic effect after administration to a subject.

[0010] Some aspects of the disclosure provide compositions comprising a Cas9 protein associated with a gRNA and a cationic lipid. It was surprisingly found that when a Cas9 protein is associated with a gRNA, the complex can be encapsulated by cationic lipids and effectively delivered to cells. This may be accomplished with or without a supercharged protein. In some embodiments, the composition comprises a Cas9 protein associated with a negatively supercharged protein (*e.g.*, supernegatively charged GFP) and a cationic lipid, which also provides for successful delivery to a cell. In some embodiments, the composition exhibits low toxicity when delivered to a population of cells, for example, wherein at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cells are viable following administration of the composition. In some embodiments, the Cas9 protein is a wild type Cas9 protein, a Cas9 nickase, or comprises a nuclease inactivated (dCas9) protein. In some embodiments, the Cas9 protein is a fusion protein comprising dCas9. In some embodiments, the fusion protein comprises a transcriptional activator (*e.g.*, VP64), a transcriptional repressor (*e.g.*, KRAB, SID) a nuclease domain (*e.g.*, FokI), a recombinase domain (*e.g.*, Hin, Gin, or Tn3), a deaminase (*e.g.*, a cytidine deaminase or an adenosine deaminase) or an epigenetic modifier domain (*e.g.*, TET1).

[0011] Other aspects of the disclosure provide compositions comprising a Cas9 protein associated with a gRNA and a cationic polymer. As with cationic lipids, when a Cas9 protein is associated with a gRNA, the complex can associate with cationic polymers and be effectively delivered to cells. This may be accomplished with or without a supercharged protein. In some embodiments, the composition comprises a Cas9 protein associated with a negatively supercharged protein (*e.g.*, supernegatively charged GFP) and a cationic polymer, which also provides for successful delivery to a cell. In some embodiments, the composition exhibits low toxicity when delivered to a population of cells, for example, wherein at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cells are viable following administration of the composition. In some embodiments, the Cas9 protein is a wild type Cas9 protein, a Cas9 nickase, or comprises a nuclease inactivated (dCas9) protein. In some embodiments, the Cas9 protein is a fusion protein comprising dCas9. In some embodiments, the fusion protein comprises a transcriptional activator (*e.g.*, VP64), a transcriptional repressor (*e.g.*, KRAB, SID) a nuclease domain (*e.g.*, FokI), a recombinase domain (*e.g.*, Hin, Gin, or Tn3), a deaminase (*e.g.*, a cytidine deaminase or an adenosine deaminase) or an epigenetic modifier domain (*e.g.*, LSD1, TET1).

[0012] Some aspects of this disclosure provide methods for administering a composition provided herein to a subject. In some embodiments, the method comprises administering a composition described herein to a subject. In some embodiments, the subject is susceptible to, is suffering from, or is displaying one or more symptoms of a disease, disorder, or condition. In some embodiments, the composition is administered to the subject in an amount sufficient and under suitable conditions for at least one sign or symptom to be ameliorated as a result of the administration. In some embodiments, the step of administering is performed under conditions sufficient for the functional effector protein to penetrate a cell of the subject. In some embodiments, the disease, disorder, or condition is associated with abnormally elevated levels of an mRNA, a protein, or combination thereof. In some embodiments, the composition comprises a nuclease that specifically binds and cleaves a genomic sequence, for example, a normal or a pathogenic allele; a gene associated with susceptibility to, or onset or progression of, a disease; a gene encoding a pathogenic RNA or protein; or a gene encoding an RNA or protein that is expressed at abnormally high levels in diseased cells or tissue. In some embodiments, the step of administering comprises a route of administration selected from the group consisting of oral, intravenous, intramuscular, intra-arterial, subcutaneous, intraventricular, topical, inhalational, and mucosal delivery.

[0013] Some aspects of this disclosure provide methods for introducing a functional effector protein into a cell. In some embodiments, the method comprises contacting the cell with a composition comprising a supercharged protein and a functional effector protein as described herein under conditions suitable for the functional effector protein to enter the cell, thereby introducing the functional effector protein into the cell. In some embodiments, the method comprises contacting the cell with a composition comprising a Cas9 protein and a cationic lipid and/or cationic polymer under conditions suitable for the Cas9 protein to enter the cell, thereby introducing the Cas9 protein into the cell. In some embodiments, the Cas9 protein enters the nucleus of the cell, for example the Cas9 protein is directed to the nucleus by including a nuclear localization signal (NLS) in the protein. In some embodiments, the method further comprises confirming that the functional effector protein (*e.g.*, including Cas9) has penetrated the cell. In some embodiments, the cell is in a subject, and the contacting is done *in vivo*. In some embodiments, the subject is diagnosed with having or being at risk of developing a disease associated with an abnormal expression level of a gene, and wherein the functional effector protein (*e.g.*, including Cas9) modulates the expression level of the gene. In some embodiments, the method further comprises detecting a change in the level of expression of the gene or detecting a therapeutic response in the subject. In some embodiments, the cell is a somatic cell. In some embodiments, the cell is contacted with the composition or the pharmaceutical composition in an amount, for a time, and under conditions sufficient to induce programming of the cell to a desired cell fate. In some embodiments, the method further comprises using the programmed cell in a cell replacement therapeutic approach. In some embodiments, the cell is a cell carrying a genomic allele associated with a disease and the functional effector protein specifically targets the allele. In some embodiments, the cell is contacted *ex vivo* and re-administered to the subject after successful targeting of the undesired allele by the functional effector protein.

[0014] Some aspects of this disclosure provide kits comprising a composition as described herein, for example, a composition comprising a supercharged protein associated with a functional effector protein. In some embodiments, the kits comprises a Cas9 protein and a supercharged protein. In some embodiments, the kits comprises a Cas9 protein and a cationic lipid. In some embodiments, the kits comprises a Cas9 protein and a cationic polymer. In some embodiments, the kit further comprises instructions for using the components included in the kit.

[0015] These and other aspects and embodiments of the invention, as well as various advantages and utilities will be more apparent with respect to the drawings and detailed description of the invention.

DEFINITIONS

[0016] As used herein and in the claims, the singular forms "a," "an," and "the" include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to "an agent" includes a single agent and a plurality of agents.

[0017] The term "associated with" as used herein in the context of two or more moieties (*e.g.*, proteins or protein domains) refers to the fact that the moieties are physically associated with or connected to one another, either directly or via one or more additional moieties that serve as a linking agent, to form a structure that is sufficiently stable so that the moieties remain physically associated under the conditions in which the structure is used, *e.g.*, under physiological conditions. A supercharged protein may be associated with a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) through non-covalent interactions (*e.g.*, electrostatic interactions). In certain embodiments, a supercharged protein may be associated with a functional effector protein through electrostatic interactions to form a complex. In some embodiments, a sufficient number of weaker interactions can provide sufficient stability for moieties to remain physically associated under a variety of different conditions. In certain embodiments, a supercharged protein is associated with a functional effector protein via a covalent bond (*e.g.*, an amide bond). In some embodiments, a functional effector protein is associated with a supercharged protein directly by a peptide bond, or indirectly via a linker.

[0018] The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active or inactive DNA cleavage domain of Cas9 or a partially inactive DNA cleavage domain (*e.g.*, a Cas9 "nickase"), and/or the gRNA binding domain of Cas9). In some embodiments, the term "Cas9" refers to a fusion protein comprising Cas9 or a fragment thereof.

[0019] In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref:

NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); or *Neisseria. meningitidis* (NCBI Ref: YP_002342100.1).

[0020] The term "cationic lipid" refers to a lipid which has a cationic, or positive, charge at physiologic pH. Cationic lipids can take a variety of forms including, but not limited to, liposomes or micelles. Cationic lipids useful for certain aspects of the present disclosure are known in the art, and, generally comprise both polar and non-polar domains, bind to polyanions, such as nucleic acid molecules or negatively supercharged proteins, and are typically known to facilitate the delivery of nucleic acids into cells. Examples of useful cationic lipids include polyethylenimine, polyamidoamine (PAMAM) starburst dendrimers, Lipofectin (a combination of DOTMA and DOPE), Lipofectase, LIPOFECTAMINE® (*e.g.*, LIPOFECTAMINE® 2000, LIPOFECTAMINE® 3000, LIPOFECTAMINE® RNAiMAX, LIPOFECTAMINE® LTX), SAINT-RED (Synvolux Therapeutics, Groningen Netherlands), DOPE, Cytofectin (Gilead Sciences, Foster City, Calif.), and Eufectins (JBL, San Luis Obispo, Calif.). Exemplary cationic liposomes can be made from N-[1-(2,3-dioleoloxy)-propyl]-N,N,N-trimethylammonium chloride (DOTMA), N-[1 -(2,3-dioleoloxy)-propyl]-N,N,N-trimethylammonium methylsulfate (DOTAP), 3β-[N-(N′,N′-dimethylaminoethane)carbamoyl]cholesterol (DC-Chol), 2,3,-dioleyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate (DOSPA), 1,2-dimyristyloxypropyl-3-dimethyl-hydroxyethyl ammonium bromide; and dimethyldioctadecylammonium bromide (DDAB). Cationic lipids have been used in the art to deliver nucleic acid molecules to cells (see, *e.g.*, U.S. Pat. Nos. 5,855,910; 5,851,548; 5,830,430; 5,780,053; 5,767,099; 8,569,256; 8,691,750; 8,748,667; 8,758,810; 8,759,104; 8,771,728; Lewis *et al.* 1996. Proc. Natl. Acad. Sci. USA 93:3176; Hope *et al.* 1998. Molecular Membrane Biology 15:1). In addition, other lipid compositions are also known in the art and include, *e.g.*, those taught in U.S. Pat. No. 4,235,871; U.S. Pat. Nos. 4,501,728; 4,837,028; 4,737,323.

[0021] The term "cationic polymer," as used herein, refers to a polymer having a net positive charge. Cationic polymers are well known in the art, and include those described in Samal *et al.*, Cationic polymers and their therapeutic potential. *Chem Soc Rev*. 2012 Nov 7;41(21):7147-94; in published U.S. patent applications U.S. 2014/0141487 A1, U.S. 2014/0141094 A1, U.S. 2014/0044793 A1, U.S. 2014/0018404 A1, U.S. 2014/0005269 A1, and U.S. 2013/0344117 A1; and in U.S. Pat. Nos. 8,709,466; 8,728,526; 8,759,103; and

8,790,664; the entire contents of each are incorporated herein by reference. Exemplary cationic polymers include, but are not limited to, polyallylamine (PAH); polyethyleneimine (PEI); poly(L-lysine) (PLL); poly(L-arginine) (PLA); polyvinylamine homo- or copolymer; a poly(vinylbenzyl-tri-C_1-C_4-alkylammonium salt); a polymer of an aliphatic or araliphatic dihalide and an aliphatic N,N,N',N'-tetra-C_1-C_4-alkyl-alkylenediamine; a poly(vinylpyridin) or poly(vinylpyridinium salt); a poly(N,N-diallyl-N,N-di-C_1-C_4-alkyl-ammoniumhalide); a homo- or copolymer of a quaternized di-C_1-C_4-alkyl-aminoethyl acrylate or methacrylate; POLYQUAD™; a polyaminoamide; and the like.

[0022] The term "deaminase" refers to an enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uracil or deoxyuracil, respectively.

[0023] The term "effective amount," as used herein, refers to an amount of a biologically active agent that is sufficient to elicit a desired biological response. For example, in some embodiments, an effective amount of a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) may refer to the amount of the protein that is sufficient to induce a detectable effect (*e.g.*, cleavage of a target site, modification of a target site, modulation of gene expression, *etc.*). Such an effect may be detected in a suitable assay, *e.g.*, in a cell-free assay, or in a target cell, tissue, or subject organism. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a functional effector protein, may vary depending on various factors as, for example, on the desired biological response, the specific allele to be targeted, the genome, target site, cell, or tissue being targeted, and the supercharged protein being used.

[0024] The term "effector protein" refers to a protein that modulates a biological function of a cell when introduced into the cell, *e.g.*, a modification of a nucleic acid molecule in the cell (such as a cleavage, deamination, recombination, *etc.*), or a modulation (*e.g.*, increases or decreases) the expression or the expression level of a gene in the cell.

[0025] The term "engineered," as used herein refers to a protein molecule, complex, substance, or entity that has been designed, produced, prepared, synthesized, and/or manufactured by a human. Accordingly, an engineered product is a product that does not occur in nature. In some embodiments, an engineered protein or composition, *e.g.*, an engineered supercharged protein associated with a functional effector protein, such as a nuclease, Cas9 protein (including variants and fusions thereof) is a supercharged protein that has been designed to meet particular requirements or to have particular desired features, *e.g.*,

to have a specified net charge, to specifically bind and/or cleave or modify a target sequence of interest, to have a specific minimal or maximal cleavage or enzymatic activity, and/or to have a specific stability.

[0026] The term "epigenetic modifier," as used herein, refers to a protein or catalytic domain thereof having enzymatic activity that results in the epigenetic modification of DNA, for example chromosomal DNA. Epigenetic modifications include, but are not limited to DNA methylation and demethylation; histone modifications including methylation and demethylation (*e.g.*, mono-, di- and tri-methylation), histone acetylation and deacetylation, as well we histone ubiquitylation, phosphorylation, and sumoylation.

[0027] The term "functional protein" refers to a protein that is in a form in which it exhibits a property and/or activity by which it is characterized.

[0028] The term "fusion protein" refers to a protein comprising a plurality of heterologous proteins, protein domains, or peptides, *e.g.*, a supercharged protein and a functional effector protein, associated with each other via a peptide linkage, thus forming a single amino acid sequence. In certain embodiments, a fusion protein is encoded by a gene.

[0029] The term "gene" has its meaning as understood in the art. It will be appreciated by those of ordinary skill in the art that the term "gene" may include gene regulatory sequences (*e.g.*, promoters, enhancers, *etc.*) and/or intron sequences. It will further be appreciated that definitions of gene include references to nucleic acids that do not encode proteins but rather encode functional RNA molecules such as RNAi agents, ribozymes, tRNAs, *etc.* For the purpose of clarity it should be noted that, as used in the present application, the term "gene" generally refers to a portion of a nucleic acid that encodes a protein; the term may optionally encompass regulatory sequences, as will be clear from context to those of ordinary skill in the art. This definition is not intended to exclude application of the term "gene" to non-protein–coding expression units but rather to clarify that, in most cases, the term as used in this document refers to a protein-coding nucleic acid.

[0030] The term "isolated" refers to a molecule, complex, substance, or entity that has been (1) separated from at least some of the components with which it was associated when initially produced (whether in nature or in an experimental setting), and/or (2) produced, prepared, synthesized, and/or manufactured by a human. Isolated substances and/or entities may be separated from at least about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, or more of the other components with which they were initially associated. In some embodiments, isolated agents are more than about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%,

about 96%, about 97%, about 98%, about 99%, or more than about 99% pure. As used herein, a substance is "pure" if it is substantially free of other components.

[0031] The term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, a supercharged protein and a nuclease. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker comprises an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is a cleavable linker, *e.g.*, the linker comprises a bond that can be cleaved upon exposure to a cleaving activity, such as UV light or a hydrolytic enzyme, such as a lysosomal protease. In some embodiments, the linker is any stretch of amino acids having at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, or more amino acids. In some embodiments, the peptide linker comprises repeats of the tri-peptide Gly-Gly-Ser, *e.g.*, comprising the sequence $(GGS)_n$, wherein n represents at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more repeats. In some embodiments, the linker comprises the sequence $(GGS)_6$ (SEQ ID NO:2). In some embodiments, the peptide linker is the 16 residue "XTEN" linker, or a variant thereof (See, *e.g.*, Schellenberger *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* **27,** 1186–1190 (2009)). In some embodiments, the XTEN linker comprises the sequence SGSETPGTSESATPES (SEQ ID NO:3), SGSETPGTSESA (SEQ ID NO:4), or SGSETPGTSESATPEGGSGGS (SEQ ID NO:5). In some embodiments, the peptide linker is one or more selected from VPFLLEPDNINGKTC (SEQ ID NO:6), GSAGSAAGSGEF (SEQ ID NO:7), SIVAQLSRPDPA (SEQ ID NO:8), MKIIEQLPSA (SEQ ID NO:9), VRHKLKRVGS (SEQ ID NO:10), GHGTGSTGSGSS (SEQ ID NO:11), MSRPDPA (SEQ ID NO:12); or GGSM (SEQ ID NO:13).

[0032] The term "nuclease," as used herein, refers to an agent, for example, a protein or a small molecule, capable of cleaving a phosphodiester bond connecting nucleotide residues in a nucleic acid molecule. In some embodiments, a nuclease is a protein, *e.g.*, an enzyme that can bind a nucleic acid molecule and cleave a phosphodiester bond connecting nucleotide residues within the nucleic acid molecule. A nuclease may be an endonuclease, cleaving a phosphodiester bond within a polynucleotide chain, or an exonuclease, cleaving a phosphodiester bond at the end of the polynucleotide chain. In some embodiments, a nuclease is a site-specific nuclease, binding and/or cleaving a specific phosphodiester bond

within a specific nucleotide sequence, which is also referred to herein as the "recognition sequence," the "nuclease target site," or the "target site." In some embodiments, a nuclease recognizes a single stranded target site, while in other embodiments, a nuclease recognizes a double-stranded target site, for example, a double-stranded DNA target site. The target sites of many naturally occurring nucleases, for example, many naturally occurring DNA restriction nucleases, are well known to those of skill in the art. In many cases, a DNA nuclease, such as EcoRI, HindIII, or BamHI, recognize a palindromic, double-stranded DNA target site of 4 to 10 base pairs in length, and cut each of the two DNA strands at a specific position within the target site. Some endonucleases cut a double-stranded nucleic acid target site symmetrically, *i.e.*, cutting both strands at the same position so that the ends comprise base-paired nucleotides, also referred to herein as blunt ends. Other endonucleases cut a double-stranded nucleic acid target site asymmetrically, *i.e.*, cutting each strand at a different position so that the ends comprise unpaired nucleotides. Unpaired nucleotides at the end of a double-stranded DNA molecule are also referred to as "overhangs," *e.g.*, as "5′-overhang" or as "3′-overhang," depending on whether the unpaired nucleotide(s) form(s) the 5′ or the 3′ end of the respective DNA strand. Double-stranded DNA molecule ends ending with unpaired nucleotide(s) are also referred to as sticky ends, as they can "stick to" other double-stranded DNA molecule ends comprising complementary unpaired nucleotide(s). A nuclease protein typically comprises a "binding domain" that mediates the interaction of the protein with the nucleic acid substrate, and a "cleavage domain" that catalyzes the cleavage of the phosphodiester bond within the nucleic acid backbone. In some embodiments, a nuclease protein can bind and cleave a nucleic acid molecule in a monomeric form, while, in other embodiments, a nuclease protein has to dimerize or multimerize in order to cleave a target nucleic acid molecule. Binding domains and cleavage domains of naturally occurring nucleases, as well as modular binding domains and cleavage domains that can be combined to create nucleases that bind specific target sites, are well known to those of skill in the art. For example, transcriptional activator like elements can be used as binding domains to specifically bind a desired target site, and fused or conjugated to a cleavage domain, for example, the cleavage domain of FokI, to create an engineered nuclease cleaving the desired target site.

[0033] The term "nucleic acid" and the term "nucleic acid molecule," as used interchangeably herein, refer to a compound comprising a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are

linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *i.e.* analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5′ to 3′ direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (*e.g.*, 2′-fluororibose, ribose, 2′-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5′-*N*-phosphoramidite linkages).

[0034] The term "pharmaceutical composition," as used herein, refers to a composition that can be administrated to a subject, for example, in the context of treatment of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, *e.g.*, a supercharged protein associated with a functional effector protein,

such as a nuclease, or a nucleic acid encoding a supercharged protein and a functional effector protein, *e.g.*, in the form of a fusion protein, and a pharmaceutically acceptable excipient.

[0035] The term "physiological pH" as used herein refers to a pH value that is found in a normal, non-pathologic cell or subject. In some embodiments, physiological pH is between pH 5 – 8. In some embodiments, physiological pH is pH 7-7.5, for example, pH 7.0, pH 7.1, pH 7.2, pH 7.3, pH 7.4, or pH 7.5. In some embodiments, physiological pH is pH 6.5-7.5. In some embodiments, physiological pH is pH 5, pH 5.5, pH 6, pH 6.5, pH 7, pH 7.5, or pH 8.

[0036] The term "prevention" or "prevent" refer to the prophylactic treatment of a subject who is at risk of developing a disease, disorder, or condition (*e.g.*, at an elevated risk as compared to a control subject, or a control group of subject, or at an elevated risk as compared to the average risk of an age-matched and/or gender-matched subject), resulting in a decrease in the probability that the subject will develop the disease, disorder, or condition (as compared to the probability without prevention), and/or to the inhibition of further advancement of an already established disorder.

[0037] The term "proliferative disease," as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasias. Malignant neoplasms are also referred to as cancers.

[0038] The term "protein" is interchangeably used herein with the terms "peptide" and "polypeptide" and refers to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any

combination thereof. A protein may comprise different domains, for example, a TALE effector protein may comprise a nucleic acid binding domain and an effector domain, *e.g.*, a nucleic acid cleavage domain or a transcriptional activator or repressor domain. In some embodiments, a protein comprises a proteinaceous part, *e.g.*, an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, *e.g.*, a compound that can act as a nucleic acid cleavage agent.

[0039] The term "RNA-programmable nuclease," and "RNA-guided nuclease" are used interchangeably herein and refer to a nuclease that forms a complex with (*e.g.*, binds or associates with) one or more RNA molecule that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. RNA-programmable nucleases include Cas9. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though "gRNA" is used interchangeably to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (*e.g.*, and directs binding of a Cas9 complex to the target); and (2) a domain that binds a Cas9 protein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site and providing the sequence specificity of the nuclease:RNA complex.

[0040] The term "recombinase," as used herein, refers to a site-specific enzyme that mediates the recombination of DNA between recombinase recognition sequences, which results in the excision, integration, inversion, or exchange (*e.g.*, translocation) of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (*e.g.*, resolvases and invertases) and tyrosine recombinases (*e.g.*, integrases). Examples of serine recombinases include, without limitation, Hin, Gin, Tn3, β-six, CinH, ParA, γδ, Bxb1, ϕC31, TP901, TG1, φBT1, R4, φRV1, φFC1, MR11, A118, U153, and gp29. Examples of tyrosine recombinases include, without limitation, Cre, FLP, R, Lambda, HK101, HK022, and pSAM2. The serine and tyrosine recombinase names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange. Recombinases have numerous applications, including the creation of gene knockouts/knock-ins and gene therapy applications. See, *e.g.*, Brown *et al.*, "Serine

recombinases as tools for genome engineering." *Methods.* 2011;53(4):372-9; Hirano *et al.*, "Site-specific recombinases as tools for heterologous gene integration." *Appl. Microbiol. Biotechnol.* 2011; 92(2):227-39; Chavez and Calos, "Therapeutic applications of the ΦC31 integrase system." *Curr. Gene Ther.* 2011;11(5):375-81; Turan and Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." *FASEB J.* 2011; 25(12):4088-107; Venken and Bellen, "Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and ΦC31 integrase." *Methods Mol. Biol.* 2012; 859:203-28; Murphy, "Phage recombinases and their applications." *Adv. Virus Res.* 2012; 83:367-414; Zhang *et al.*, "Conditional gene manipulation: Cre-ating a new biological era." *J. Zhejiang Univ. Sci. B.* 2012; 13(7):511-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." *DNA Repair (Amst).* 2012; 1;11(10):781-8; the entire contents of each are hereby incorporated by reference in their entirety. The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in embodiments of the invention. The methods and compositions of the invention can be expanded by mining databases for new orthogonal recombinases or designing synthetic recombinases with defined DNA specificities (See, *e.g.*, Groth *et al.*, "Phage integrases: biology and applications." *J. Mol. Biol.* 2004; 335, 667-678; Gordley *et al.*, "Synthesis of programmable integrases." *Proc. Natl. Acad. Sci. U S A.* 2009; 106, 5053-5058; the entire contents of each are hereby incorporated by reference in their entirety). Other examples of recombinases that are useful in the methods and compositions described herein are known to those of skill in the art, and any new recombinase that is discovered or generated is expected to be able to be used in the different embodiments of the invention. In some embodiments, a recombinase (or catyltic domain thereof) is fused to a Cas9 protein (*e.g.*, dCas9).

[0041] The term "recombine," or "recombination," in the context of a nucleic acid modification (*e.g.*, a genomic modification), is used to refer to the process by which two or more nucleic acid molecules, or two or more regions of a single nucleic acid molecule, are modified by the action of a recombinase protein. Recombination can result in, *inter alia*, the insertion, inversion, excision, or translocation of a nucleic acid sequence, *e.g.*, in or between one or more nucleic acid molecules.

[0042] The term "subject," as used herein, refers to an individual organism. In some embodiments, the subject is a human of either sex at any stage of development. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments,

the subject is a laboratory animal, for example, a mouse, a rat, a gerbil, a guinea pig, a fish, a frog, or a fly. In some embodiments, the subject is a farm animal, for example, a sheep, a goat, a pig, or a cattle. In some embodiments, the subject is a companion animal, for example, a cat or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is genetically engineered, *e.g.*, a genetically engineered non-human subject.

[0043] The term "supercharge" refers to any modification of a protein that results in the increase or decrease of the overall net charge of the protein. Modifications include, but are not limited to, alterations in amino acid sequence or addition of charged moieties (*e.g.*, carboxylic acid groups, phosphate groups, sulfate groups, amino groups). Supercharging also refers to the association of an agent with a charged protein, naturally occurring or modified, to form a complex with increased or decreased charge relative to the agent alone.

[0044] The term "target site," as used herein in the context of functional effector proteins that bind a nucleic acid molecule, such as nucleases and transcriptional activators or repressors, refers to a sequence within a nucleic acid molecule that is bound and acted upon by the effector protein, *e.g.*, cleaved by the nuclease or transcriptionally activated or repressed by the transcriptional activator or repressor, respectively. A target site may be single-stranded or double-stranded. In the context of RNA-guided (*e.g.*, RNA-programmable) nucleases (*e.g.*, Cas9), a target site typically comprises a nucleotide sequence that is complementary to the gRNA of the RNA-programmable nuclease, and a protospacer adjacent motif (PAM) at the 3′ end adjacent to the gRNA-complementary sequence. For the RNA-guided nuclease Cas9 (or variants or fusions comprising having gRNA binding activity), the target site may be, in some embodiments, 20 base pairs plus a 3 base pair PAM (*e.g.*, NNN, wherein N represents any nucleotide). Typically, the first nucleotide of a PAM can be any nucleotide, while the two downstream nucleotides are specified depending on the specific RNA-guided nuclease. Exemplary target sites for RNA-guided nucleases, such as Cas9, are known to those of skill in the art and include, without limitation, NNG, NGN, NAG, and NGG, wherein N represents any nucleotide. In addition, Cas9 nucleases from different species (*e.g.*, *S. thermophilus* instead of *S. pyogenes*) recognizes a PAM that comprises the sequence NGGNG. Additional PAM sequences are known, including, but not limited to, NNAGAAW and NAAR (see, *e.g.*, Esvelt and Wang, *Molecular Systems Biology*, 9:641 (2013), the entire contents of which are incorporated herein by reference). For example, the target site of an RNA-guided nuclease, such as, *e.g.*, Cas9, may comprise the structure [NZ]-[PAM], where each N is, independently, any nucleotide, and Z is an integer

between 1 and 50, inclusive. In some embodiments, Z is at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, or at least 50. In some embodiments, Z is 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, or 50. In some embodiments, Z is 20. In some embodiments, "target site" may also refer to a sequence within a nucleic acid molecule that is bound but not cleaved by a nuclease. For example, certain embodiments described herein provide proteins comprising an inactive (or inactivated) Cas9 DNA cleavage domain. Such proteins (*e.g.*, when also including a Cas9 RNA binding domain) are able to bind the target site specified by the gRNA, however because the DNA cleavage site is inactivated, the target site is not cleaved by the particular protein. However, such proteins as described herein are typically conjugated, fused, or bound by another protein (*e.g.*, a nuclease, transcriptional activator, recombinase, deaminase, *etc.*) or molecule that mediates modification of the nucleic acid molecule. In some embodiments, the sequence actually cleaved will depend on the protein (*e.g.*, nuclease) or molecule that mediates cleavage of the nucleic acid molecule, and in some cases, for example, will relate to the proximity or distance from which the inactivated Cas9 protein(s) is/are bound. In the context of nucleases that dimerize, for example, dimers of a protein comprising an inactive Cas9 (or a Cas9 RNA binding domain) and a DNA cleavage domain (*e.g.*, FokI cleavage domain or an active Cas9 cleavage domain), a target sites typically comprises a left-half site (bound by one protein), a right-half site (bound by the second protein), and a spacer sequence between the half sites in which the cut is made. In some embodiments, either the left-half site or the right half-site (and not the spacer sequence) is cut. This structure ([left-half site]-[spacer sequence]-[right-half site]) is referred to herein as an LSR structure. In some embodiments, the left-half site and/or the right-half site correspond to an RNA-guided target site (*e.g.*, a Cas9 target site). In some embodiments, either or both half-sites are shorter or longer than, *e.g.*, a typical region targeted by Cas9, for example shorter or longer than 20 nucleotides. In some embodiments, the left and right half sites comprise different nucleic acid sequences. In some embodiments, the target site is a sequence comprising three (3) RNA-guided nuclease target site sequences, for example, three sequences corresponding to Cas9 target site sequences, in which the first and second, and second and third Cas9 target site sequences are separated by a spacer sequence. In some embodiments, the spacer sequence is at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least

14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 125, at least 150, at least 175, at least 200, or at least 250 bp long.

[0045] The terms "transcriptional activator" and "transcriptional repressor" refer to an agent such as a protein (*e.g.*, a transcription factor or fragment thereof), that binds a target nucleic acid sequence and causes an increase or decrease of the level of expression of a gene product associated with the target nucleic acid sequence, respectively. For example, if the target nucleic acid sequence is located within a regulatory region of a gene, a transcriptional activator causes an increase of the level of expression of a gene product encoded by the gene (conversely, a transcriptional repressor causes a decrease of the level of expression of a gene product encoded by the gene). The gene product can be an RNA transcribed from the gene (*e.g.*, an mRNA) or a polypeptide translated from an mRNA transcribed from the gene. Typically an increase or decrease in the level of an mRNA results in an or decrease increase in the level of a polypeptide translated therefrom. The level of expression may be determined using standard techniques for measuring mRNA or protein.

[0046] The term "Transcriptional Activator-Like Effector," (TALE) as used herein, refers to effector proteins comprising a DNA binding domain, which contains a highly conserved 33-34 amino acid sequence comprising a highly variable two-amino acid motif (Repeat Variable Diresidue, RVD). The RVD motif determines binding specificity to a nucleic acid sequence, and can be engineered according to methods well known to those of skill in the art to specifically bind a desired DNA sequence (see, *e.g.*, Miller, Jeffrey; *et.al.* (February 2011). "A TALE nuclease architecture for efficient genome editing". *Nature Biotechnology* **29** (2): 143–8; Zhang, Feng; *et.al.* (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". *Nature Biotechnology* **29** (2): 149–53; Geiβler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". *PLoS ONE* **6** (5): e19509; Boch, Jens (February 2011). "TALEs of genome targeting". *Nature Biotechnology* **29** (2): 135–6; Boch, Jens; *et.al.* (December 2009). "Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors". *Science* **326** (5959): 1509–12; and Moscou, Matthew J.; Adam J. Bogdanove (December 2009). "A Simple Cipher Governs DNA Recognition by TAL Effectors". *Science* **326** (5959): 1501; the entire contents of each of which are incorporated herein by reference). The simple relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA binding domains by selecting a

combination of repeat segments containing the appropriate RVDs. TALE effector proteins include, without limitation, TALE nucleases (TALENs) and TALE transcriptional activators and repressors.

[0047] The term "Transcriptional Activator-Like Element Nuclease," (TALEN) as used herein, refers to an artificial nuclease comprising a transcriptional activator like effector DNA binding domain to a DNA cleavage domain, for example, a FokI domain. A number of modular assembly schemes for generating engineered TALE constructs have been reported (Zhang, Feng; *et.al.* (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". *Nature Biotechnology* **29** (2): 149–53; Geiβler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". *PLoS ONE* **6** (5): e19509; Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. *et al.* (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting". *Nucleic Acids Research*; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). "Assembly of custom TALE-type DNA binding domains by modular cloning". *Nucleic Acids Research*; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes". *Nucleic Acids Research.*; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). Bendahmane, Mohammed. ed. "Assembly of Designer TAL Effectors by Golden Gate Cloning". *PLoS ONE* **6** (5): e19722; each of which is incorporated herein by reference).

[0048] The term "transcriptional repressor" refers to a transcription factor, *e.g.*, a protein, that binds a target nucleic acid sequence and causes a reduction of the level of expression of a gene product associated with the target nucleic acid sequence. For example, if the target nucleic acid sequence is located within a regulatory region of a gene, a transcriptional repressor causes a reduction of the level of expression of a gene product encoded by the gene. The gene product can be an RNA transcribed from the gene (*e.g.*, an mRNA) or a polypeptide translated from an mRNA transcribed from the gene. Typically a reduction in the level of an mRNA results in a reduction in the level of a polypeptide translated therefrom. The level of expression may be determined using standard techniques for measuring mRNA or protein.

[0049] The term "zinc finger nuclease," as used herein, refers to a nuclease comprising a nucleic acid cleavage domain conjugated to a binding domain that comprises a

zinc finger array. In some embodiments, the cleavage domain is the cleavage domain of the type II restriction endonuclease FokI. Zinc finger nucleases can be designed to target virtually any desired sequence in a given nucleic acid molecule for cleavage, and the possibility to design zinc finger binding domains to bind unique sites in the context of complex genomes allows for targeted cleavage of a single genomic site in living cells, for example, to achieve a targeted genomic alteration of therapeutic value. Targeting a double-strand break to a desired genomic locus can be used to introduce frame-shift mutations into the coding sequence of a gene due to the error-prone nature of the non-homologous DNA repair pathway. Zinc finger nucleases can be generated to target a site of interest by methods well known to those of skill in the art. For example, zinc finger binding domains with a desired specificity can be designed by combining individual zinc finger motifs of known specificity. The structure of the zinc finger protein Zif268 bound to DNA has informed much of the work in this field and the concept of obtaining zinc fingers for each of the 64 possible base pair triplets and then mixing and matching these modular zinc fingers to design proteins with any desired sequence specificity has been described (Pavletich NP, Pabo CO (May 1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A". *Science* **252** (5007): 809–17, the entire contents of which are incorporated herein). In some embodiments, separate zinc fingers that each recognizes a 3 base pair DNA sequence are combined to generate 3-, 4-, 5-, or 6-finger arrays that recognize target sites ranging from 9 base pairs to 18 base pairs in length. In some embodiments, longer arrays are contemplated. In other embodiments, 2-finger modules recognizing 6-8 nucleotides are combined to generate 4-, 6-, or 8- zinc finger arrays. In some embodiments, bacterial or phage display is employed to develop a zinc finger domain that recognizes a desired nucleic acid sequence, for example, a desired nuclease target site of 3-30 bp in length. Zinc finger nucleases, in some embodiments, comprise a zinc finger binding domain and a cleavage domain fused or otherwise conjugated to each other via a linker, for example, a polypeptide linker. The length of the linker determines the distance of the cut from the nucleic acid sequence bound by the zinc finger domain. If a shorter linker is used, the cleavage domain will cut the nucleic acid closer to the bound nucleic acid sequence, while a longer linker will result in a greater distance between the cut and the bound nucleic acid sequence. In some embodiments, the cleavage domain of a zinc finger nuclease has to dimerize in order to cut a bound nucleic acid. In some such embodiments, the dimer is a heterodimer of two monomers, each of which comprise a different zinc finger binding domain. For example, in some embodiments, the dimer may comprise one monomer comprising zinc finger domain A conjugated to a FokI

cleavage domain, and one monomer comprising zinc finger domain B conjugated to a FokI cleavage domain. In this non-limiting example, zinc finger domain A binds a nucleic acid sequence on one side of the target site, zinc finger domain B binds a nucleic acid sequence on the other side of the target site, and the dimerize FokI domain cuts the nucleic acid in between the zinc finger domain binding sites.

[0050] The term "zinc finger," as used herein, refers to a small nucleic acid-binding protein structural motif characterized by a fold and the coordination of one or more zinc ions that stabilize the fold. Zinc fingers encompass a wide variety of differing protein structures (see, *e.g.*, Klug A, Rhodes D (1987). "Zinc fingers: a novel protein fold for nucleic acid recognition". Cold Spring Harb. Symp. Quant. Biol. 52: 473–82, the entire contents of which are incorporated herein by reference). Zinc fingers can be designed to bind a specific sequence of nucleotides, and zinc finger arrays comprising fusions of a series of zinc fingers, can be designed to bind virtually any desired target sequence. Such zinc finger arrays can form a binding domain of a protein, for example, of a nuclease, *e.g.*, if conjugated to a nucleic acid cleavage domain. Different types of zinc finger motifs are known to those of skill in the art, including, but not limited to, Cys_2His_2, Gag knuckle, Treble clef, Zinc ribbon, Zn_2/Cys_6, and TAZ2 domain-like motifs (see, *e.g.*, Krishna SS, Majumdar I, Grishin NV (January 2003). "Structural classification of zinc fingers: survey and summary". *Nucleic Acids Res.* 31 (2): 532–50). Typically, a single zinc finger motif binds 3 or 4 nucleotides of a nucleic acid molecule. Accordingly, a zinc finger domain comprising 2 zinc finger motifs may bind 6-8 nucleotides, a zinc finger domain comprising 3 zinc finger motifs may bind 9-12 nucleotides, a zinc finger domain comprising 4 zinc finger motifs may bind 12-16 nucleotides, and so forth. Any suitable protein engineering technique can be employed to alter the DNA-binding specificity of zinc fingers and/or design novel zinc finger fusions to bind virtually any desired target sequence from 3 – 30 nucleotides in length (see, *e.g.*, Pabo CO, Peisach E, Grant RA (2001). "Design and selection of novel cys2His2 Zinc finger proteins". *Annual Review of Biochemistry* 70: 313–340; Jamieson AC, Miller JC, Pabo CO (2003). "Drug discovery with engineered zinc-finger proteins". *Nature Reviews Drug Discovery* 2 (5): 361–368; and Liu Q, Segal DJ, Ghiara JB, Barbas CF (May 1997). "Design of polydactyl zinc-finger proteins for unique addressing within complex genomes". *Proc. Natl. Acad. Sci. U.S.A.* 94 (11); the entire contents of each of which are incorporated herein by reference). Fusions between engineered zinc finger arrays and protein domains that cleave a nucleic acid can be used to generate a "zinc finger nuclease." A zinc finger nuclease typically comprises a zinc finger domain that binds a specific target site within a nucleic acid

molecule, and a nucleic acid cleavage domain that cuts the nucleic acid molecule within or in proximity to the target site bound by the binding domain. Typical engineered zinc finger nucleases comprise a binding domain having between 3 and 6 individual zinc finger motifs and binding target sites ranging from 9 base pairs to 18 base pairs in length. Longer target sites are particularly attractive in situations where it is desired to bind and cleave a target site that is unique in a given genome.

[0051] The terms "treatment," "treat," and "treating," refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms "treatment," "treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, *e.g.*, to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (*e.g.*, in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

BRIEF DESCRIPTION OF THE DRAWINGS

[0052] *Figure 1.* Schematic of macromolecular delivery into mammalian cells.

[0053] *Figure 2.* Programming adipocyte cell fate: the switch from White Adipose Tissue (WAT) to Brown Adipose Tissue (BAT).

[0054] *Figure 3.* Using supercharged delivery platforms to deliver TALE activators programmed to target PPARγ or PRDM16.

[0055] *Figure 4.* Schematic of a fusion protein comprising a +36 GFP fusion, an 18.5 mer TALE domain, and a VP64 activation domain.

[0056] *Figure 5.* Expression and purification of the +36 GFP-TALE activator-fusion protein.

[0057] *Figure 6.* Testing for activation of fat cell regulator genes upon delivery of +36 GFP PPARγ and PRDM16 TALE activator fusion proteins.

[0058] *Figure 7.* Delivery efficacy of +36 GFP TALE activator fusion proteins at different concentrations.

[0059] *Figure 8.* Comparison of delivery efficacy of two different +36 GFP-PRDM16 TALE fusion proteins in NIH 3T3 cells.

[0060] *Figure 9.* PPARγ gene expression after delivery of PPARγ-TALE activator fusion and comparison to various controls.

[0061] *Figure 10.* PRDM16 gene expression after delivery of RDM16-TALE activator fusion and comparison to various controls.

[0062] *Figure 11.* Moderate TALE activity is observed in the presence of serum.

[0063] *Figure 12.* Validation of viral delivery of PPARγ followed by 7-day treatment with adipogenesis cocktail.

[0064] *Figure 13.* Schematic of an assay for programming fibroblasts into WAT and BAT.

[0065] *Figure 14.* Adipocyte formation observed upon treatment with +36 GFP TALE activator fusion protein.

[0066] *Figure 15.* Staining of various treatments after 7 days with LipidTOX red shows formation of adipocytes after viral delivery as well as after delivery of supercharged PPARγ TALE activator fusion protein.

[0067] *Figure 16.* Staining of various treatments after 7 days with LipidTOX red shows formation of adipocytes after viral delivery as well as after delivery of supercharged PPARγ TALE activator fusion protein.

[0068] *Figure 17.* Expression of WAT biomarker genes after viral delivery as well as after delivery of supercharged PPARγ TALE activator fusion protein.

[0069] *Figure 18.* Delivery of supercharged PRDM16 TALE activator fusion proteins to induce brown-fat adipocytes in vivo. Robust adipocyte formation was observed after viral delivery of PPARγ and PRDM16 and also after delivery of supercharged TALE activator protein fusions.

[0070] *Figure 19.* Comparison of TALE/TALE, viral/TALE, and viral/viral-induced expression of brown fat markers by expression of PPARγ and PRDM16.

[0071] *Figure 20.* RT-qPCR assessments are consistent with fat cell differentiation observed by LipidTOX staining.

[0072] *Figure 21.* Delivery of functional TALE activator fusion proteins as complexes with +36 GFP improves TALE activator activity after delivery.

[0073] *Figure 22.* PRDM16 gene expression after TALE activator fusion delivery either as a fusion (+36GFP PRDM16 TALE-3) or a complex (+36GFP + PRDM16 TALE-3) with +36GFP. Delivery of complexes tends to increase TALE activator activity.

[0074] *Figure 23.* Effect of Aurein peptide fusion to +36GFP on PRDM16 gene expression after TALE activator fusion delivery (either as a fusion or a complex with +36GFP).

[0075] *Figure 24.* PRDM16 gene expression after TALE activator fusion delivery either as a fusion (+36GFP PRDM16 TALE-3) or a complex (+36GFP + PRDM16 TALE-3) with Lipofectamine LTX.

[0076] *Figure 25.* Delivery of supercharged fusion proteins or complexes with Cas9 into mammalian cells. (GGS)9-T-ALAL-PKKKRKV corresponds to SEQ ID NO: 251.

[0077] *Figure 26.* Purification of wild-type Cas9 protein and Cas9 fusion proteins with +36GFP and Aurein-GGS9.

[0078] *Figure 27A-B.* A strategy for delivering proteins into mammalian cells by fusion or non-covalent complexation with polyanionic macromolecules and encapsulation with cationic lipids is shown. Figure 27(A) shows that recombinases, transcriptional-activator-like effector (TALE) proteins, and Cas9 endonucleases bind nucleic acids and are natively cationic (net theoretical charges are shown in black) and are not efficiently encapsulated by cationic lipids. These proteins can be rendered highly anionic, however, by fusion to either a supernegatively charged protein such as (–30)GFP, or by complexation with polyanionic nucleic acids. Figure 27(B) shows a schematic representing that cationic lipids commonly used to transfect DNA and RNA encapsulate the resulting highly anionic proteins or protein:nucleic acid complexes, mediating their delivery into mammalian cells.

[0079] *Figure 28A-F.* Delivery of Cre recombinase to cultured human cells. In Figure 28(A), the fusion of either highly cationic (+36)GFP or highly anionic (–30)GFP to Cre recombinase is shown. A HeLa reporter cell line that expresses DsRed upon Cre-mediated recombination was used to evaluate Cre delivery efficiency; (GGS)9 corresponds to SEQ ID NO : 252 and His6 corresponds to SEQ ID NO: 253. In Figure 28(B) HeLa dsRed cells treated with 10 nM (–30)GFP-Cre and the cationic lipid RNAiMAX. Cells were visualized after incubation for 48 hours in media containing 10% fetal bovine serum (FBS). Figure 28(C) shows the delivery of (+36)GFP-Cre in 10% FBS media or in serum-free media, and (–30)GFP-Cre with or without the cationic lipid RNAiMAX in full-serum media. Figure 28(D) presents the effect of cationic lipid dose on functional (–30)GFP-Cre delivery efficacy after 48 hours in 275 μL media containing 10% FBS. Figure 28(E) is a comparison of several commercially available cationic lipids and polymers for functional delivery efficacy of (–30)dGFP-Cre. Figure 28(F) shows the RNAiMAX-mediated delivery of multiple anionic

peptide or protein sequences fused to Cre. Error bars reflect the standard deviation from three biological replicates performed on different days.

[0080] *Figure 29A-B.* Delivery of TALE transcriptional activators into cultured human cells. Figure 29(A) shows the design of an 18.5-repeat TALE activator fused C-terminally to a VP64 activation domain and N-terminally to (–30)GFP and an NLS; (GGS)9 corresponds to SEQ ID NO : 251 and His6 corresponds to SEQ ID NO: 252. The overall net theoretical charge of the fusion is –43. Figure 29(B) demonstrates the activation of NTF3 transcription by traditional transfection of plasmids encoding TALE-VP64 activators that target sites in the NTF3 gene, or by RNAiMAX cationic lipid-mediated delivery of the corresponding NTF3-targeting (–30)GFP-TALE-VP64 proteins. Gene expression levels were measured by qRT-PCR and are normalized to GAPDH expression levels. Error bars reflect the standard deviation from three biological replicates performed on different days.

[0081] *Figure 30A-E.* Delivery of Cas9:sgRNA, Cas9 D10A nickase, and dCas9-VP64 transcriptional activators to cultured human cells. Figure 30(A) demonstrates the cationic lipid-mediated delivery of Cas9 protein variants complexed with an EGFP-targeting sgRNA or a VEGF-targeting sgRNA to U2OS EGFP reporter cells. Results are compared to that of standard transfection of Cas9 and sgRNA expression plasmids. Figure 30(B) shows the results of a T7 endonuclease I (T7EI) assay to measure the modification of EGFP from no treatment (lane 1), treatment with EGFP-targeting sgRNA alone (lane 2), Cas9 protein alone (lane 3), Cas9 protein + VEGF-targeting sgRNA + RNAiMAX (lane 4), transfection of plasmids expressing Cas9 and EGFP-targeting sgRNA (lane 5), or Cas9 protein + EGFP-targeting sgRNA + RNAiMAX (lane 6). Indel efficiencies calculated by densitometry are shown below the gel image. Figure 30(C) presents the results of a T7EI assay of genome modification at EGFP and three endogenous genes with a single delivery of Cas9 complexed with four sgRNAs and RNAiMAX. Indel efficiencies calculated by densitometry are shown below the gel image. Figure 30(D) shows the delivery of Cas9 D10A nickase and pairs of sgRNAs either by plasmid transfection or by RNAiMAX-mediated protein:RNA complex delivery. EGFP-disrupting sgRNAs GFP g1 + GFP g5, or GFP g3 + GFP g7, are expected to result in gene disruption, while GFP g5 + GFP g7 target the same strand and are therefore expected to be non-functional. Figure 30(E) shows the delivery of catalytically dead (dCas9)-VP64 transcriptional activators that target NTF3 either by plasmid transfection or RNAiMAX-mediated protein delivery. Delivery of both VEGF g3 and VEGF g5 sgRNAs served as a negative control for NTF3 gene activation. Error bars reflect the standard deviation from six biological replicates performed on different days.

[0082] *Figure 31A-B.* The DNA sequence specificity of Cas9-mediated endogenous gene cleavage in cultured human cells by plasmid transfection or by cationic lipid-mediated protein:sgRNA delivery is shown. In Figure 31(A), a T7EI assay was performed for on-target modification of endogenous CLTA, EMX, and VEGF genes. In Figure 31(B-D) the on-target:off-target DNA modification ratio resulting from Cas9:sgRNA for plasmid transfection or cationic lipid-mediated protein:sgRNA delivery is shown. The conditions for each treatment were adjusted to result in ~10% on-target cleavage, enabling a comparison of DNA cleavage specificity between the two delivery methods under conditions in which on-target gene modification efficiencies are comparable. P values are listed in Table 2. Each on- and off-target sample was sequenced once with > 10,000 sequences analyzed per on-target sample and an average of > 111,000 sequences analyzed per off-target sample (Table 2).

[0083] *Figures 32A-D.* The *in vivo* delivery of Cre recombinase and Cas9:sgRNA complexes to hair cells in the mouse inner ear is shown. In Figure 32(A), the scala media (cochlear duct) of P0 floxP-tdTomato mice (n = 4) were injected with 0.3 μL of 23 μM (–30)GFP-Cre in 50% RNAiMAX or with RNAiMAX alone (control). After 5 days, tdTomato expression indicative of Cre-mediated recombination was visualized using immunohistology. Red = tdTomato; green = Myo7a; white = Sox2; blue = DAPI. Yellow brackets indicate the outer hair cell (OHC) layer. Figure 32(B) shows that, ten days after (–30)GFP-Cre delivery, intact espin (Esp)-expressing stereocilia of tdTomato-positive outer hair cells were present (arrow), similar to stereocilia in control cochlea. Red = tdTomato; green = Esp; white = Sox2; blue = DAPI. Figure 32(C) is identical to Figure 32(A) except using Lipofectamine 2000 instead of RNAiMAX. (n = 4). The upper and lower panels are images of mice cochlea at low and high magnification, respectively, detailing the efficiency of delivery as well as the effect on cochlear architecture and hair cell loss. Figure 32(D) shows the results when the scala media (cochlear duct) of P2 Atoh1-GFP mice (n = 3) were injected with 0.3 μL of 33 μM Cas9, 33 μM sgRNA in 50% RNAiMAX or Lipofectamine 2000. Cas9-mediated gene disruption results in the loss of GFP expression when visualized 10 days later. The upper panels show GFP signal only, while lower panels include additional immunohistological markers. Yellow boxes in the lower panels highlight hair cells that have lost GFP expression. Red = tdTomato; green = Myo7a; white/light blue = Sox2; blue = DAPI. All scale bars, shown in white, are 10 μm.

[0084] *Figure 33A-C.* Optimization of cationic lipid-mediated delivery of Cre recombinase. Figure 33(A) shows the optimization of (–30)GFP-Cre delivery in BSR-TdTomato cells, a second reporter cell line used for measuring Cre recombination efficiency.

Figure 33B demonstrates the effect of RNAiMAX dosage on (–30)GFP-Cre recombination efficiency in HeLa dsRed reporter cells and toxicity as measured by FACS. HeLa cells were sorted by forward-scatter and side-scatter gating to identify live cells that retained normal morphology. Figure 33(C) illustrates the relationship between net charge of the protein fused to Cre recombinase and cationic lipid-mediated functional Cre delivery efficiency. Cre recombinase fused to the domains listed at 25 nM were combined with 1.5 μL RNAiMAX and incubated with HeLa dsRed reporter cells. After 2 days, recombination efficiency was measured by FACS. Error bars reflect the standard deviation from three biological replicates performed on different days.

[0085] *Figures 34A-D.* Protein uptake by cationic lipid-mediated delivery versus superpositively charged cationic protein delivery. Figure 34(A) quantifies GFP fluorescence from cells treated with either (–30)GFP-Cre and RNAiMAX or (+36)GFP-Cre after washing cells with PBS + heparin (20 U/mL) to remove unbound protein. Figure 34(B) shows the functional Cre recombinase delivery efficiency of (–30)GFP-Cre + 1.5 μL RNAiMAX relative to Cre recombinase delivery efficiency arising from fusion with (+36)GFP. Figure 34(C) provides a comparison of mCherry uptake by (–30)GFP-fusion + 1.5 μM RNAiMAX treatment versus (+36)GFP fusion by measuring mean mCherry fluorescence of total cell population 48 hours after treatment and washing cells with PBS + heparin. Figure 34(D) shows the total cellular GFP fluorescence of (–30)GFP-Cre or (+36)GFP-Cre in the presence or absence of RNAiMAX.

[0086] *Figure 35.* Delivery optimization of TALE activators designed to target the *NTF3* gene. HEK293T cells were treated with either NTF3 TALE plasmid by transfection of by liposomal delivery of NTF3 TALE proteins. Cells were harvested after empirically determined optimal incubation time for both treatments and analyzed by qRT-PCR for mRNA levels of NTF3. Optimal protein (25-50 nM) and lipid dosage (1.5 μL RNAiMAX) was chosen for comparison of two delivery techniques in Figure 29B. Error bars reflect the standard deviation from six biological replicates performed on different days.

[0087] *Figures 36A-D.* Determination of gene disruption frequency of an *EGFP* reporter gene by delivery of Cas9:sgRNA and analyzing by flow cytometry. Figure 36(A) provides a schematic of EGFP disruption in U2OS cells by NHEJ induced by Cas9 double-stranded breaks. Figure 36(B) shows the delivery of *EGFP*-targeting sgRNA or an off-target sgRNA complexed with (–30)dGFP-Cas9 using RNAiMAX along with a plasmid transfection positive control (orange). Figure 36(C) provides confirmation that disruption of EGFP fluorescence is not a result of cellular toxicity by treating samples with the TO-PRO-3

live/dead stain (Life Technologies, Carlsbad CA) and analyzing the resulting cells by flow cytometry. Figure 36(D) shows testing of the TO-PRO-3 stain by addition of a cell permeabilizing, but not completely membrane lysing, detergent (0.5% Tween).

[0088] *Figures 37A-D.* Optimization of Cas9:sgRNA functional delivery. In Figure 37(A), cationic lipid-mediated delivery efficiency of two tested constructs shows that the more anionic (–30)dGFP-NLS-Cas9 facilitates more efficient delivery at low protein and sgRNA concentrations compared with native Cas9. Figure 37(B) shows the delivery optimization of (–30)dGFP-NLS-Cas9 as a function of protein and sgRNA concentration. Figure 37(C) shows the delivery of Cas9 protein without any fusions or tags as a function of protein and sgRNA concentration. Figure 37(D) provides the optimal sgRNA to protein ratio for RNAiMAX-mediated delivery of (–30)dGFP-NLS-Cas9 and native Cas9. Error bars reflect standard deviation from three biological replicates performed on different days.

[0089] *Figures 38A-C.* The effect of the NLS and/or (–30)dGFP on functional Cas9 delivery as a function of both sgRNA and Cas9 concentration. *EGFP* gene disruption was measured at three fixed sgRNA concentrations: 10 nM (Figure 38(A)), 25 nM (Figure 38(B)), and 50 nM (Figure 38(C)), along with varying protein concentrations show in the graphs. Delivery was performed using 0.8 μL RNAiMAX and assayed by FACS 48 hours later for loss of EGFP fluorescence signal.

[0090] *Figures 39A-C.* Effects of RNAiMAX and Lipofectamine 2000 on Cas9:sgRNA delivery efficiency and cellular toxicity. In Figure 39(A), *EGFP* gene disruption at different Cas9 protein concentrations and a constant dose of 100 nM EGFP sgRNA in U2OS EGFP reporter cells treated for 16 hours with either RNAiMAX or Lipofectamine 2000 is shown. After 16 hours, media was removed and fresh media was added to cells until end point of assay 48-72 hours post protein delivery treatment. The live cell population was determined by FACS using TO-PRO-3 Live/Dead stain. Figure 39(B) shows the toxicity profile for Cas9:sgRNA delivery to U2OS cells as a function of Lipofectamine 2000 dose. Figure 39(C) provides the toxicity profile for cells as a function of RNAiMAX dose. Error bars reflect standard deviation from three biological replicates performed on different days.

[0091] *Figure 40.* Optimization of dCas9-VP64 delivery targeting the *NTF3* gene at varying concentrations of protein and sgRNA. HEK293T cells were treated with dCas9-VP64 activator and either *NTF3*–targeting gRNA g2 or a mixture of all six *NTF3*–targeting sgRNAs for 16 hours and 0.8 μL RNAiMAX in 48-well plate format (275 μL final volume). NTF3 mRNA levels were determined by qRT-PCR and normalized to those of GAPDH.

Error bars reflect standard deviation from six biological replicates performed on different days.

[0092] *Figures 41A-C.* Indel frequencies, measured by high-throughput sequencing, of several human genes treated either by a mock treatment, by transfection of Cas9 plasmid and sgRNA linear DNA PCR product, or by cationic lipid-mediated protein:sgRNA delivery are depicted. Mock treatment involved cationic lipid-mediated protein:sgRNA delivery of *EGFP*-targeting sgRNA instead of one of the three human gene-targeting sgRNAs. Figure 41(A) shows the on-target and off-target indel frequencies for the *CLTA* gene. Figure 41(B) provides the on-target and off-target indel frequencies for the *EMX* gene. Figure 41(C) demonstrates the on-target and off-target indel frequencies for the *VEGF* gene. Each on- and off-target sample was sequenced once with > 10,000 sequences analyzed per on-target sample and an average of > 111,000 sequences analyzed per off-target sample (Table 2).

[0093] *Figures 42A-C.* Delivery of Cas9 endonuclease to mouse embryonic stem cells. Figure 42(A) shows floating spheres treated with Cas9 protein and RNAiMAX but no sgRNA (control) retained strong GFP fluorescence (right), while those treated with Cas9:sgRNA and RNAiMAX exhibited decreased GFP fluorescence (left). Scale bars are 100 μm. Figure 42(B) shows the control progenitor cells after cell attachment, and virtualy all the control progenitor cells were GFP positive (right panels). Cas9:sgRNA treatment led to significant reduction in GFP expression (left panels) and many progenitor cells showed complete GFP knockdown (arrows) after cell attachment. Scale bars are 20 μm. Figure 42(C) shows a T7EI assay on stem cells harvested after imaging confirm cleavage of GFP reporter. Similar gene target modification efficiencies were observed from cationic lipid-mediated Cas9:sgRNA delivery and transfection of Cas9 and EGFP sgRNA plasmids.

[0094] *Figures 43A-B.* Genome modification induced by cationic lipid-mediated protein delivery of Cas9 nuclease and sgRNA at endogenous loci *in vivo*. Figure 43(A) shows representative examples of genomic DNA sequences at the *EGFP* on-target locus that are modified following cationic lipid-mediated delivery of Cas9 and EGFP sgRNA in mouse hair cells. For each example shown, the unmodified genomic site is the first sequence, followed by the most abundant eight sequences containing deletions and three sequences containing insertions. The numbers before each sequence indicate sequencing counts. The sgRNA target sites are bold and underlined in green. Insertions and deletions are shown in red. PAM site is shown in blue. Figure 43(B) shows an identical analysis as in Figure 42(A) for *EMX* on-target site in mouse hair cells. The sequences shown in Figure 43(A), from top to bottom,

correspond to SEQ ID NOs:223-236; and the sequences shown in Figure 43(B), from top to bottom, correspond to SEQ ID NOs:237-250.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

[0095] The present invention provides complexes, compositions, preparations, kits, systems, and related methods for the delivery of functional effector proteins, *e.g.*, nucleases, recombinases, and Cas9 proteins (including variants and fusions thereof, *e.g.*, Cas9 nickases and Cas9 fusions to deaminases, gene editing enzymes, transcriptional repressors and activators, epigenetic modifiers, *etc.*), to a cell by associating the functional effector protein with one or more of a supercharged protein, cationic polymer, and/or cationic lipid. Typically, the functional effector protein is delivered to the interior of a cell, *e.g.*, to cause a biological effect in the cell, such as cleavage of a genomic target sequence or modulation of the expression of a target gene. In some embodiments, the biological effect exerts a therapeutic benefit to a subject in which the cell is found. The complexes, compositions, preparations, systems, kits, and related methods for delivery of functional effector proteins are useful for introducing an effector protein into a cell, *e.g.*, in the context of manipulating the cell for a research or therapeutic purpose. The compositions, preparations, systems, kits, and related methods for delivery of functional effector proteins provided herein exhibit improved efficacy and reduced cytotoxicity, and ease of preparation as compared to current technologies. The delivery of site-specific proteins, such as TALENs or Cas9 proteins (or variants or fusions thereof) using the compositions, preparations, systems, kits, and related methods provided herein allows for the targeted manipulation/modification of the genome of a host cell *in vitro* or *in vivo* while avoiding the use of more invasive delivery methods, such as viral delivery of vectors encoding site-specific proteins.

[0096] In some embodiments, the inventive technology uses a supercharged protein to deliver a functional effector protein into a cell. In certain embodiments, the supercharged protein is an engineered protein. In some embodiments, the supercharged protein is a naturally occurring supercharged protein. Some aspects of this invention are based on the recognition that supercharged proteins are endocytosed by cells; that functional effector proteins that can be associated with supercharged proteins are effectively taken up by cells together with the supercharged proteins; and that such functional effector proteins retain their biological function after cellular uptake, *e.g.*, in that they are able to cleave or modify genomic target sites or modulate transcription of a target gene.

[0097] In some embodiments, the compositions provided herein comprising a supercharged protein associated with a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) are useful as therapeutic agents, diagnostic agents, or research tools. In some embodiments, a functional effector protein, such as a nuclease or a transcription factor, may be therapeutically active, *e.g.*, in that it targets a gene associated with a disease or disorder. In some embodiments, a composition as provided herein, comprising a supercharged protein and a functional effector protein, such as a nuclease or a transcription factor, is used to modulate the expression of a gene in a cell or to modulate a biological pathway (*e.g.*, a signaling pathway, a metabolic pathway) in a cell. In some embodiments, a cell is contacted with an inventive composition described herein to introduce a functional effector protein into the cell. In some embodiments, an inventive composition is administered to a subject in need thereof to introduce a functional effector protein into a cell within the subject, *e.g.*, into a cell associated with a disease or disorder. Suitable cells and cell types for delivery of functional effector proteins according to some aspects of this disclosure include, but are not limited to, human cells, mammalian cells, T-cells, neurons, stem cells, progenitor cells, blood cells, fibroblasts, epithelial cells, neoplastic cells, and tumor cells.

Supercharged Proteins

[0098] Supercharged proteins for use in the present invention can be produced by changing non-conserved amino acids on the surface of a protein to more polar or charged amino acid residues. In certain embodiments, non-conserved amino acids on the surface of the protein are mutated into amino acids that are positively charged at physiological pH (pH ~7.4). The amino acid residues to be modified may be hydrophobic, hydrophilic, charged, or a combination thereof. Supercharged proteins can also be produced by the attachment of charged moieties to the protein in order to supercharge the protein. Supercharged proteins frequently are resistant to aggregation, have an increased ability to refold, resist improper folding, have improved solubility, and are generally more stable under a wide range of conditions, including denaturing conditions such as heat or the presence of a detergent.

[0099] Supercharged proteins suitable for use according to aspects of this disclosure are known in the art and include, without limitation, those supercharged proteins disclosed in international PCT patent application, PCT/US07/70254, filed June 1, 2007, published as WO 2007/143574 on December 13, 2007; in international PCT application, PCT/US09/041984,

filed on April 28, 2009, published as WO 2009/134808 on November 5, 2009; and in international PCT application, PCT/US10/001250, filed on April 28, 2010, published as WO 2010/129023 on November 11, 2010; the entire contents of each of which are incorporated herein by reference. In some embodiments, the supercharged protein is an engineered supercharged protein. In some embodiments, the supercharged protein is a naturally occurring supercharged protein, *e.g.*, a naturally supercharged protein disclosed in international PCT application, PCT/US10/001250, filed on April 28, 2010, published as WO 2010/129023 on November 11, 2010; each of which is incorporated herein by reference. In some embodiments, the supercharged protein, engineered or naturally occurring, exhibits a charge:molecular weight ratio of greater than 0.8, *e.g.*, ≥0.85, ≥0.9, ≥0.95, ≥1, ≥1.1, ≥1.2, ≥1.3, ≥1.4, ≥1.5, ≥1.6, ≥1.7, ≥1.8, ≥1.9, ≥2, ≥2.5, ≥3, ≥4, ≥5, ≥6, ≥7, ≥8, or ≥10.

[00100] The supercharged protein employed may be derived from any species of plant, animal, and/or microorganism. In certain embodiments, the supercharged protein is a mammalian protein. In certain embodiments, the supercharged protein is a human protein. In certain embodiments, the protein is derived from an organism typically used in research. For example, the protein to be modified may be from a primate (*e.g.*, ape, monkey), rodent (*e.g.*, rabbit, hamster, gerbil), pig, dog, cat, fish (*e.g.*, *Danio rerio*), nematode (*e.g.*, *C. elegans*), yeast (*e.g.*, *Saccharomyces cerevisiae*), or bacteria (*e.g.*, *E. coli*). In certain embodiments, the protein is non-immunogenic. In certain embodiments, the protein is non-antigenic. In certain embodiments, the protein does not have inherent biological activity or has been modified to have no biological activity. In certain embodiments, the protein is chosen based on its targeting ability. In certain embodiments, the protein is a green fluorescent protein. In some embodiments, the supercharged protein is supercharged glutathione S-transferace (GST). In some embodiments, the supercharged protein is supercharged streptavidin.

[00101] In some embodiments, a supercharged protein is used that has been modified to increase the overall net charge, or to increase the total number of charged residues on the protein surface. In certain embodiments, the theoretical net charge of the supercharged protein is increased by at least +1, at least +2, at least +3, at least +4, at least +5, at least +10, at least +15, at least +20, at least +25, at least +30, at least +35, or at least +40 as compared to the unmodified protein. In certain embodiments, the theoretical net charge of the supercharged protein is at least +1, at least +2, at least +3, at least +4, at least +5, at least +10, at least +15, at least +20, at least +25, at least +30, at least +35, or at least +40 at physiological pH (*i.e.*, ~7.4).

[00102] In other embodiments, for example those involving use of cationic lipids and/or cationic polymers, a supercharged protein is used that has been modified to decrease the overall net charge, or to decrease the total number of charged residues on the protein surface. In certain embodiments, the theoretical net charge of the supercharged protein is decreased ("minus" or "negative" represented by '-') by at least -1, at least -2, at least -3, at least -4, at least -5, at least -10, at least -15, at least -20, at least -25, at least -30, at least -35, at least -40, at least -45, or at least -50 as compared to the unmodified protein. In certain embodiments, the theoretical net charge of the supercharged protein is at least -1, at least -2, at least -3, at least -4, at least -5, at least -10, at least -15, at least -20, at least -25, at least -30, at least -35, at least -40, at least -45, or at least -50.

[00103] While some exemplary supercharged proteins are described herein in order to exemplify the inventive technology, the disclosure is not limited in this respect. Those of skill in the art will be able to ascertain additional suitable supercharged proteins for delivering functional effector proteins to cells based on the instant disclosure. A number of naturally occurring proteins may be modified to generate suitable supercharged proteins. The desired modifications in such proteins may be accomplished using any techniques known in the art. Recombinant DNA techniques for introducing such changes in a protein sequence are well known in the art. In certain embodiments, the modifications are made by site-directed mutagenesis of the polynucleotide encoding the protein. Other techniques for introducing mutations are discussed in *Molecular Cloning: A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch, and Maniatis (Cold Spring Harbor Laboratory Press: 1989); the treatise, *Methods in Enzymology* (Academic Press, Inc., N.Y.); Ausubel *et al.*, *Current Protocols in Molecular Biology* (John Wiley & Sons, Inc., New York, 1999); each of which is incorporated herein by reference.

[00104] Supercharged proteins may be further modified. Proteins including supercharged proteins can be modified using techniques known to those of skill in the art. For example, supercharged proteins may be modified chemically or biologically. One or more amino acids may be added, deleted, or changed from the primary sequence. For example, a poly-histidine tag or other tag may be added to the supercharged protein to aid in the purification of the protein. Other peptides or proteins may be added onto the supercharged protein to alter the biological, biochemical, and/or biophysical properties of the protein. For example, an endosomolytic peptide may be added to the primary sequence of the supercharged protein, or a targeting peptide, may be added to the primary sequence of the supercharged protein. Other modifications of the supercharged protein include, but are not

limited to, post-translational modifications (*e.g.*, glycosylation, phosphorylation, acylation, lipidation, farnesylation, acetylation, proteolysis, *etc.*). In certain embodiments, the supercharged protein is modified to reduce its immunogenicity. In certain embodiments, the supercharged protein is modified to enhance its ability to deliver a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) to a cell. In certain embodiments, the supercharged protein is conjugated to a polymer. For example, the protein may be PEGylated by conjugating the protein to a polyethylene glycol (PEG) polymer. Other methods can be used to produce supercharged proteins without modification of the protein sequence. For example, moieties that alter the net charge can be attached to proteins (*e.g.*, by chemical or enzymatic reactions) to provide surface charge to achieve supercharging. In certain embodiments, the method of modifying proteins described in Shaw *et al.*, *Protein Science* 17:1446, 2008 is used to supercharge a protein that is used in the instantly disclosed inventive technology.

[00105] The design and creation of variants of several different supercharged proteins suitable for use with the instantly disclosed technology is described in international PCT patent application, PCT/US07/70254, filed June 1, 2007, published as WO 2007/143574 on December 13, 2007; in international PCT application, PCT/US09/041984, filed on April 28, 2009, published as WO 2009/134808 on November 5, 2009; and in international PCT application PCT/US10/001250, filed on April 28, 2010, published as WO 2010/129023 on November 11, 2010; the entire contents of each of which are incorporated herein by reference. Some of the disclosed supercharged proteins described therein have been shown to be more stable and to retain their biological function, *e.g.*, their fluorescence in the case of fluorescent proteins. For example, a green fluorescent protein (GFP) from *Aequorea victoria* is described in GenBank Accession Number P42212, incorporated herein by reference. The amino acid sequence of this wild type GFP is as follows:

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYP
DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVT
AAGITHGMDELYK (SEQ ID NO:14)

[00106] Wild type GFP has a theoretical net charge of -7. Variants with a theoretical net charge of -29, -30, -25, +15, +25, +36, +48, and +49 have been reported, *e.g.*, in in international PCT application PCT/US10/001250, filed on April 28, 2010, published as WO 2010/129023 on November 11, 2010, the entire contents of which are incorporated herein by reference. Even after heating the +36 GFP to 95 °C, 100% of the variant protein is soluble and the protein retains ≥70% of its fluorescence.

[00107] Some aspects of this disclosure are based on the discovery that +36 GFP efficiently delivers functional effector proteins (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) to target cells, and that the effector proteins so delivered retain their biological function. Therefore, GFP or other proteins with a net charge of at least +15, at least +25, at least +30, at least +35, or at least +40 are thought to be particularly useful for introducing functional effector proteins into a cell.

[00108] In some embodiments, particularly useful supercharged proteins are proteins that allow for a charge distribution or a surface charge density similar to that of +36 GFP. Further, in some embodiments, particularly useful supercharged proteins are proteins exhibiting a stable folded structure not easily perturbed by supercharging, thus allowing the supercharged protein to be well folded. In some embodiments, particularly useful supercharged proteins are proteins sharing a structural feature with a supercharged protein described herein or in international PCT patent application, PCT/US07/70254, filed June 1, 2007, published as WO 2007/143574 on December 13, 2007; in international PCT application, PCT/US09/041984, filed on April 28, 2009, published as WO 2009/134808 on November 5, 2009; and in international PCT application, PCT/US10/001250, filed on April 28, 2010, published as WO 2010/129023 on November 11, 2010; the entire contents of each of which are incorporated herein by reference; for example, a globular structure, or a β-barrel structure. Protein folding, protein fold structure stability and perturbation of protein folding by substitution of specific amino acids with differently charged amino acids, charge distribution, and surface charge density can be modeled *in silico* by methods and algorithms provided herein and others known to those of skill in the art. Accordingly, it will be apparent to those of skill in the art from no more than routine experimentation, whether a supercharged protein in question will be well folded. Thus, those of skill in the art will be able to identify from a given amino acid sequence whether a given supercharged protein will be useful for cellular delivery of a functional effector protein according to the technology described herein.

[00109] Some exemplary, suitable variants of GFP include, without limitation:

[00110] +15 GFP:

MGHHHHHHGGASKGERLFTGVVPILVELDGDVNGHKFSVRGEGEGDATRGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPKHMKRHDFFKSAMPEGYVQERTISFKKDGTYKTRAEVKFEGRTLVNRIELKGRDFKEKGNILGHK
LEYNFNSHNVYITADKRKNGIKANFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSALSKDPKEKR
DHMVLLEFVTAAGITHGMDELYK (SEQ ID NO:15)

[00111] +25 GFP:

MGHHHHHHGGASKGERLFTGVVPILVELDGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGTYKTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHK

LRYNFNSHNVYITADKRKNGIKANFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSALSKDPKEKR
DHMVLLEFVTAAGITHGMDELYK (SEQ ID NO:16)

[00112] +36 GFP:
MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHK
LRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKR
DHMVLLEFVTAAGIKHGRDERYK (SEQ ID NO:17)

[00113] +42 GFP:
MGHHHHHHGGRSKGKRLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHK
LRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRKHYLSTRSKLSKDPKEKR
DHMVLLEFVTAAGIKHGRKERYK (SEQ ID NO:18)

[00114] +48 GFP:
MGHHHHHHGGRSKGKRLFRGKVPILVKLKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFKGRTLVNRIKLKGRDFKEKGNILGHK
LRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLAKHYQQNTPIGRGPVLLPRKHYLSTRSKLSKDPKEKR
DHMVLLEFVTAAGIKHGRKERYK (SEQ ID NO:19)

[00115] +49 GFP:
MGHHHHHHGGRSKGKRLFRGKVPILVKLKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFKGRTLVNRIKLKGRDFKEKGNILGHK
LRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLAKHYQQNTPIGRGPVLLPRKHYLSTRSKLSKDPKEKR
DHMVLKEFVTAAGIKHGRKERYK (SEQ ID NO:20)

[00116] (-)30 GFP:
MGHHHHHHGGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLVTTLT
YGVQCFSDYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK
LEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLSTESALSKDPNEDR
DHMVLLEFVTAAGIDHGMDELYK (SEQ ID NO:21)

[00117] It will be apparent to the skilled artisan that the sequences above include an N-terminal His6 tag, and that sequences without such a tag or with a different tag are also suitable.

[00118] In order to promote the biological function of the functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) after delivery to a cell, it may be desirable to enhance endosomal escape of the functional effector protein after cellular uptake. A supercharged protein or a functional effector protein may be fused to or associated with a protein, peptide, or other entity known to enhance endosome degradation or lysis of the endosome. In certain embodiments, the peptide is hemagglutinin 2 (HA2) peptide which is known to enhance endosome degradation. In certain particular embodiments, HA2 peptide is fused to supercharged GFP (*e.g.*, +36 GFP). In certain particular embodiments, the fused protein is of the sequence:

[00119] +36 GFP-HA2
MGHHHHHHGGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHK
LRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKR
DHMVLLEFVTAAGIKHGRDERYKGSAGSAAGSGEFGLFGAIAGFIENGWEGMIDG (SEQ ID NO:22)

[00120] In certain embodiments, the endosomolytic peptide is melittin peptide (GIGAVLKVLTTGLPALISWIKRKRQQ, SEQ ID NO: 23) (Meyer *et al.*, *JACS* 130(11): 3272-3273, 2008; which is incorporated herein by reference). In certain embodiments, the melittin peptide is modified by one, two, three, four, or five amino acid substitutions, deletions, and/or additions. In certain embodiments, the melittin peptide is of the sequence: CIGAVLKVLTTGLPALISWIKRKRQQ (SEQ ID NO:24). In certain particular embodiments, the melittin peptide is fused to supercharged GFP (*e.g.*, +36 GFP).

[00121] In certain embodiments, the endosomolytic peptide is penetratin peptide (RQIKIWFQNRRMKWKK-amide, SEQ ID NO:25), bovine PrP (1-30) peptide (MVKSKIGSWILVLFVAMWSDVGLCKKRPKP-amide, SEQ ID NO: 26), MPGΔ^{NLS} peptide (which lacks a functional nuclear localization sequence because of a K->S substitution) (GALFLGWLGAAGSTMGAPKSKRKV, SEQ ID NO:27), TP-10 peptide (AGYLLGKINLKALAALAKKIL-amide, SEQ ID NO:28), and/or EB1 peptide (LIRLWSHLIHIWFQNRRLKWKKK-amide, SEQ ID NO:29) (Lundberg *et al.*, 2007, *FASEB J.* 21:2664; incorporated herein by reference). In certain embodiments, the penetratin, PrP (1-30), MPG, TP-10, and/or EB1 peptide is modified by one, two, three, four, or five amino acid substitutions, deletions, and/or additions. In certain particular embodiments, the PrP (1-30), MPG, TP-10, and/or EB1 peptide is fused to supercharged GFP (*e.g.*, +36 GFP). In some embodiments, an Aurein peptide is fused to the supercharged protein.

[00122] Other peptides or proteins may also be fused to the supercharged protein or to a fusion protein comprising a supercharged protein and a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*). For example, a targeting peptide may be fused to the supercharged protein in order to selectively deliver a functional effector protein to a particular cell type. Peptides or proteins that enhance cellular uptake of the functional effector protein may also be used. In certain embodiments, the peptide fused to the supercharged protein is a peptide hormone. In certain embodiments, the peptide fused to the supercharged protein is a peptide ligand.

[00123] The exemplary supercharged proteins described in detail herein are not meant

to limit the disclosure, and one of skill in the art will appreciate that other supercharged proteins may be used for the cellular delivery of functional effector proteins (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*), including, but not limited to, other GFP-style fluorescent proteins. In certain embodiments, the supercharged protein is a supercharged version of blue fluorescent protein. In certain embodiments, the supercharged protein is a supercharged version of cyan fluorescent protein. In certain embodiments, the supercharged protein is a supercharged version of yellow fluorescent protein. Exemplary suitable fluorescent proteins include, but are not limited to, enhanced green fluorescent protein (EGFP), AcGFP, TurboGFP, Emerald, Azami Green, ZsGreen, EBFP, Sapphire, T-Sapphire, ECFP, mCFP, Cerulean, CyPet, AmCyan1, Midori-Ishi Cyan, mTFP1 (Teal), enhanced yellow fluorescent protein (EYFP), Topaz, Venus, mCitrine, YPet, PhiYFP, ZsYellow1, mBanana, Kusabira Orange, mOrange, dTomato, dTomato-Tandem, DsRed, DsRed2, DsRed-Express (T1), DsRed-Monomer, mTangerine, mStrawberry, AsRed2, mRFP1, JRed, mCherry, HcRed1, mRaspberry, HcRed1, HcRed-Tandem, mPlum, and AQ143.

[00124] Yet other proteins that may be supercharged and used, *e.g.*, in the delivery of functional effector proteins as disclosed herein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*), include histone components or histone-like proteins, high-mobility-group proteins (HMGs), enzymes (*e.g.*, amylases, pectinases, hydrolases, proteases, glucose isomerase, lipases, phytases, alglucerase, imiglucerase, agalsidase beta, α-1-iduronidase, acid α-glucosidase, and iduronate-2-sulfatase, *N*-acetylgalactosamine-4-sulfatase.

[00125] Charged polymers other than proteins may also be used to deliver functional effector proteins. Additionally, as described in greater detail herein, cationic lipids and lipid-like materials as well as cationic polymers can also be used to deliver functional effector proteins. Suitable cationic lipids, lipid-like materials and cationic polymers are disclosed herein and additional suitable lipids and lipid-like matierals are known to those of skill in the art (see, *e.g.*, those described in Akinc *et al.*, *Nature Biotechnology* 26, 561 - 569 (2008), the entire contents of which are incorporated herein by reference).

Delivery of functional effector proteins using supercharged proteins

[0064] The present invention provides systems and methods for the delivery of functional effector proteins (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) to cells *in vivo*, *ex vivo*, or *in vitro*. Such

systems and methods typically involve association of the functional effector protein with a supercharged protein to form a complex or a fusion protein, and delivery of the complex or fusion protein to a cell. In some embodiments, the functional effector protein to be delivered by the supercharged protein has therapeutic activity. In some embodiments, delivery of the complex or fusion protein to a cell involves administering the complex or fusion protein comprising a supercharged protein associated with a functional effector protein to a subject in need thereof.

[0065] In some embodiments, a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) by itself may not be able to enter a cell, but is able to enter a cell when associated with a supercharged protein, for example, via a covalent bond or a non-covalent interaction. In some embodiments, a composition is provided that includes a functional effector protein that is covalently bound to a supercharged protein. In some embodiments, the composition includes a functional effector protein fused to a supercharged protein via a peptide bond, for example, via direct fusion or via a peptide linker. In some embodiments, the composition includes a functional effector protein that is bound to a supercharged protein by non-covalent interaction. In some embodiments, a supercharged protein is utilized to allow a functional effector protein to enter a cell. In some embodiments, the functional effector protein delivered to the cell associated with a supercharged protein is separated from the supercharged protein after delivery to the cell, for example, by cleavage of a linker peptide by a cellular protease (*e.g.*, an endosomal protease) or by dissociation of the functional effector protein from the supercharged protein in a specific cellular microenvironment, for example, in the endosome. In some embodiments, functional effector proteins delivered to a cell by a system or method provided by this disclosure have therapeutic activity.

[0066] In some embodiments, a functional effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) is delivered to a cell *in vivo*, *ex vivo*, or *in vitro* by a system, composition, or method provided herein. In some embodiments, a functional effector protein is a protein able to carry out a biological function within the target cell, for example, an enzyme able to bind its substrate and to catalyze an enzymatic reaction in the target cell, *e.g.*, a nuclease able to bind and cut a nucleic acid molecule within a target cell, or a transcription factor able to interact with the genome of a target cell and to activate or inhibit transcription of a target gene in the cell.

[0067] In some embodiments, a method for generating a fusion of a functional effector protein and a supercharged protein includes the generation of an expression nucleic acid construct containing the coding sequences of the functional protein and the supercharged protein, as well as, optionally, a peptide linker, in frame; the expression of such a recombinant fusion protein in a prokaryotic or eukaryotic cell in culture, the extraction and purification of the fusion protein of the fusion protein. In some embodiments, a nucleic acid construct is generated in the form of an expression vector, for example, a vector suitable for propagation in a bacterial host and for expression in a prokaryotic or eukaryotic cell.

[0068] In some embodiments, a vector suitable for fusion protein expression is generated by cloning of a nucleotide sequence coding for a functional effector protein to be delivered into a cloning vector including a nucleotide sequence coding for a supercharged protein under the control of a eukaryotic and/or a prokaryotic promoter, by a cloning approach that results in both coding sequences being in frame with each other. In some embodiments, the cloning vector includes a nucleotide sequence coding for a peptide linker between a nucleotide sequence coding for a supercharged protein and a restriction site useful for inserting a nucleotide sequence coding for a protein in frame with the linker and the supercharged protein. In some embodiments, the cloning vector further includes an additional sequence enhancing expression of a fusion protein in a prokaryotic or eukaryotic cell or facilitating purification of expressed fusion proteins from such cells, for example, a sequence stabilizing a transcript encoding the fusion protein, such as a poly-A signal, a spliceable intron, a sequence encoding an in-frame peptide or protein domain tag (*e.g.*, an Arg-tag, calmodulin-binding peptide tag, cellulose-binding domain tag, DsbA tag, c-myc-tag, glutathione S-transferase tag, FLAG-tag, HAT-tag, His-tag, maltose-binding protein tag, NusA tag, S-tag, SBP-tag, Strep-tag, or thioredoxin tag), or a selection marker or reporter cassette allowing for identification of cells harboring and expressing the expression construct and/or quantifying the level of expression in such cells. Methods for cloning and expressing fusion proteins are well known to those in the art, see, for example Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Volume 1-3, CSHL Press (1989); Gellissen *et al.*, Production of recombinant proteins, Wiley-VCH, 2005.

[00126] In some embodiments, the functional effector protein is associated with a supercharged GFP, for example, +36 GFP or -30 GFP, for delivery to a target cell. The benefit of endosomal disruption in the delivery of macromolecules by supercharged proteins has been previously demonstrated (Wadia *et al.*, *Nat. Med.* 10, 310-315, 2004) and in some embodiments, additional steps to effect enhanced endosomal escape, as provided

herein or known in the art, are performed. Highly efficient protein internalization, when coupled with effective endosomal release, has the potential to minimize the requisite doses of exogenous protein agents, enhancing their potential as research tools and leads for therapeutic development.

[00127] In some embodiments, a composition comprising a functional effector protein associated with a supercharged protein is administered to a target cell after isolation and/or purification. Protein isolation methods and technologies are well known to those of skill in the art and include, for example, affinity chromatography or immunoprecipitation. The methods suitable for isolating and/or purifying a specific functional effector proteins, supercharged proteins, and/or fusion proteins will depend on the nature of the respective protein. For example, a His-tagged fusion protein can readily be isolated and purified via Ni or Co ion chromatography, while fusion proteins tagged with other peptides or domains or untagged fusion proteins can be purified by other well established methods.

[00128] Functional effector proteins suitable for delivery to a target cell *in vivo*, *ex vivo*, or *in vitro*, by a system or method provided herein will be apparent to those of skill in the art and include, for example, DNA-binding proteins, such as transcription factors and nucleases, as well as Cas9 proteins (including variants and fusions thereof).

[00129] In some embodiments, a method, composition, or system provided herein is used to deliver a therapeutic functional effector protein to a cell. Examples of therapeutic proteins include, but are not limited to, nucleases and Cas9 proteins (including variants and fusions thereof) targeting a genomic allele associated with a disease or disorder, and transcription factors activating a beneficial gene or repressing a pathogenic gene.

[00130] In some embodiments, Cas9 is fused to a supercharged protein for delivery to a cell. In some embodiments, the supercharged protein is positively charged. In some embodiments, the supercharged protein fused to Cas9 is (+36)GFP. In some embodiments, the fusion of Cas9 and (+36)GFP comprises the amino acid sequence of SEQ ID NO:30 (*e.g.*, with or without a nuclear localization signal (NLS) and with or without a 6xHis tag), or comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO:30 (*e.g.*, with or without a nuclear localization signal (NLS) and with or without a 6xHis tag). In some embodiments, the supercharged protein fused to Cas9 is (-30)GFP. In some embodiments, the fusion of Cas9 and (-30)GFP comprises the amino acid sequence of SEQ ID NO:31 (*e.g.*, with or without a nuclear localization signal (NLS) and with or without a 6xHis tag), or comprises an amino acid sequence that is at least 80%, at least 85%, at least

90%, at least 95%, at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO:31 (*e.g.*, with or without a nuclear localization signal (NLS) and with or without a 6xHis tag).

[00131] (+36)dGFP-NLS-Cas9-6xHis **(Y67S)**:

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLT**S**
GVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKG
NILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLS
TRSKLSKDPKEKRDHMVLLEFVTAAGIKHGRDERYKTGGSGGSGGSGGSGGSGGSGGSGGTALAL
PKKKRKVMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD
FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV
EKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT
GLYETRIDLSQLGGD<u>HHHHHH</u> (SEQ ID NO:30)

[00132] (–30)dGFP-NLS-Cas9-6xHis **(Y67S)**:

MGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLVTTLT**S**
GVQCFSDYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDG
NILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLS
TESALSKDPNEDRDHMVLLEFVTAAGIDHGMDELYKTGGSGGSGGSGGSGGSGGSGGSGGTALAL
PKKKRKVMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD
FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY

KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFA
TVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV
EKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT
GLYETRIDLSQLGGDHHHHHH (SEQ ID NO:31)

Compositions of functional effector proteins and cationic lipids

[00133] Certain aspects of the disclosure relate to the use of cationic lipids for the delivery of effector proteins (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*), for example as opposed to delivering "naked" protein preparations. Surprisingly, existing liposomal delivery reagents that have been engineered for the delivery of nucleic acids such as DNA and RNA were found to effectively deliver certain effector proteins (*e.g.*, Cas9 proteins including variants and fusions thereof) both *in vitro* and *in vivo*, as described herein. Nucleic acid delivery has benefited greatly from the development of liposomal reagents over the past two decades. Cationic liposomal formulations have enabled DNA and RNA transfection to become a routine technique in basic research and have even been used in clinical trials. The lipid bilayer of the liposome protects encapsulated nucleic acids from degradation and can prevent specific neutralization by antibodies that can bind naked preparations of the nucleic acids. Importantly, fusion of the liposome with the endosomal membrane during endosomal maturation can enable highly efficient endosomal escape of cationic lipid-delivered cargo. Other non-cationic, but reversibly ionizable, lipid nanoparticle formulations have enabled efficient encapsulation and delivery of nucleic acids, while avoiding non-specific electrostatic interactions and consequent sequestration. However, proteins are chemically diverse, and therefore unlike highly anionic nucleic acids, liposomal formulations have not been similarly successful for the efficient delivery of proteins. For example, while proteins can be encapsulated non-specifically and delivered by rehydrated lipids *in vitro*, the efficacy of encapsulation is dependent on protein concentration and is generally inefficient, and thus has not seen widespread application. Aspects of the present disclosure relate to the recognition that anionic proteins or protein complexes (including those proteins associated with nucleic acids) may be able to take advantage of the same electrostatics-driven encapsulation used by cationic liposomal reagents for nucleic acid delivery. While few proteins natively possess the density of negative charges found in the phosphate backbone of nucleic acids, translational fusion to, or non-covalent association with, an anionic carrier such as a negatively supercharged protein or a nucleic acid as described herein render the resulting effector

protein or protein complex sufficiently anionic to drive efficient encapsulation of such protein cargoes by cationic liposomal reagents.

[00134] In some embodiments, association or fusion with an engineered supernegatively charged GFP is capable of driving efficient encapsulation and delivery of proteins into cultured mammalian cells by cationic lipids commonly used to transfect nucleic acids. This approach is effective even at low nanomolar protein concentrations and in the presence of serum, resulting in up to 1,000-fold more efficient functional protein delivery than protein delivery methods that use fusion to cationic peptides or proteins. As shown in the Examples, the efficacy of delivery depends, in some embodiments, on *e.g.*, the theoretical net charge of the fusion tag, and that popular natively anionic peptide tags *e.g.*, 3xFLAG and VP64, can likewise enable liposomal protein delivery.

[00135] The Examples further show that Cas9 nuclease protein associated with polyanionic guide RNAs (gRNA) can be efficiently delivered in functional form into mammalian cells by these common cationic liposomal formulations because, while not wishing to be bound by any particular theory, it is believed that the gRNA acts as a polyanionic mediator between the otherwise cationic Cas9 protein and the cationic lipids. Delivery of Cas9:gRNA complexes is not only highly efficient (*e.g.*, up to 80% modification from a single treatment) but also results in markedly higher genome modification specificity compared with plasmid transfection, typically resulting in >10-fold higher on-target:off-target modification ratios, presumably due to the transient nature of the delivered Cas9:gRNA activity. In some embodiments, delivery of Cas9:gRNA complexes results in at least a 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 20-fold or 25-fold or higher on-target:off-target modification ratio. The Examples also demonstrate that this protein delivery approach can be effective *in vivo*, for example by delivering functional Cre recombinase and functional Cas9:gRNA complexes to hair cells in the inner ear of mice.

[00136] Accordingly, some aspects of the disclosure provide compositions comprising a Cas9 protein (*e.g.*, as described herein; see *e.g.*, *Cas9 effector proteins* below) and a cationic lipid capable of delivering the Cas9 protein to the interior of a cell. In some embodiments, the Cas9 protein is associated with a gRNA, which *e.g.*, provides anionic charge to the complex thereby allowing the Cas9:gRNA complex to be encapsulated by the cationic lipids. In some embodiments, the Cas9 protein need not be associated with a gRNA for effective encapsulation by a cationic lipid, but instead is associated with a negatively supercharged protein, as described herein. In some embodiments where a Cas9 protein is

associated with a negatively supercharged protein, the Cas9 protein is also associated with a gRNA. In some embodiments, the Cas9 protein is a wild type Cas9 protein, a fragment of a wild type Cas9 protein, or a variant of a wild type Cas9 protein. In some embodiments, the Cas9 protein comprises a dCas9 domain (*e.g.*, as described herein). In some embodiments, the Cas9 protein is a fusion protein comprising a dCas9 domain (*e.g.*, as described herein). In some embodiments, the Cas9 protein is a Cas9 nickase.

[00137] In other embodiments, compositions comprising an effector protein (*e.g.*, other than a Cas9 protein) and a cationic lipid are provided which are capable of delivering the effector protein to the interior of a cell (*e.g.*, to the nucleus of the cell). The effector protein is either naturally negatively charged, is modified to have a net overall negative charge, or is associated with a negatively supercharged protein, as described herein. In some embodiments, the effector protein is any effector protein described herein. In some embodiments, the effector protein is a recombinase, *e.g.*, any recombinase described herein. In some embodiments, the recombinase is Cre recombinase. In some embodiments, the Cre recombinase comprises the amino acid sequence of SEQ ID NO:32 (*e.g.*, with or without the 6xHis tag). In some embodiments, the Cre recombinase comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO:32 (*e.g.*, with or without the 6xHis tag). In some embodiments, the Cre recombinase is fused to a supercharged protein (*e.g.*, +36 GFP or -30GFP). In some embodiments, the Cre recombinase fused to a supercharged protein comprises the amino acid sequence of SEQ ID NO:33 (*e.g.*, with or without the 6xHis tag) or SEQ ID NO:34 (*e.g.*, with or without the 6xHis tag), or comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO:33 or SEQ ID NO:34 (*e.g.*, with or without the 6xHis tag). In some embodiments, the effector protein is a TALE protein, (*e.g.*, as described herein including those provided in the Examples). In some embodiments, the TALE protein comprises one or more of a VP64 transcriptional activator domain (*e.g.*, SEQ ID NO:35). In some embodiments, the TALE protein with a VP64 transcriptional activator domain further comprises an amino acid sequence selected from the group consisting of SEQ ID NO:36-39 (*e.g.*, with or without the 6xHis tag). In some embodiments, the TALE protein with a VP64 transcriptional activator domain comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:36-39 (*e.g.*, with or without the 6xHis tag). In some embodiments, the TALE effector protein comprises a (-

30)GFP domain (*e.g.*, SEQ ID NO:21 or SEQ ID NO:40), a N-terminal region of a TALE domain (*e.g.*, SEQ ID NO:41), a variable repeat domain (*e.g.*, an 18.5mer repeat domain as provided in Maeder et al., "Robust, synergistic regulation of human gene expression using TALE activators." *Nat. Methods.* 2013; **10,** 243–245), a C-terminal TALE domain (*e.g.*, SEQ ID NO:42), a VP64 activation domain (*e.g.*, SEQ ID NO:35), and optionally one or more linkers (*e.g.*, GGS(9), SEQ ID NO: 252) between any domain and optionally a sequence tag (*e.g.*, 6xHis, SEQ ID NO:253).

[00138] While liposomal delivery of cargo such as DNA and RNA has been known to induce toxicity in targeted cells, it was found that the inventive compositions described herein deliver their cargo both *in vitro* and *in vivo* surprisingly with no or low toxicity. For example, in some embodiments, the compositions comprising a Cas9 protein or other effector proteins described herein exhibit low toxicity when administered to a population of cells (*e.g.*, *in vitro* or *in vivo*). In some embodiments, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cells in a population are viable following administration of an inventive composition comprising a Cas9 protein or other effector protein and cationic lipids. Methods for assessing the toxicity of a composition when administered to a population of cells are well known in the art and include those described in the Examples.

[00139] Cre-6xHis (6xHis tag underlined):

MASNLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAE
PEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQAL
AFERTDFDQVRSLMENSDRCQDIRNLAFLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTL
VSTAGVEKALSLGVTKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRL
IYGAKDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLL
EDGDGGSHHHHHH (SEQ ID NO:32)

[00140] (+36)GFP-Cre-6xHis (+36 GFP double-underlined; 6xHis tag underlined):

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLTY
GVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKG
NILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLS
TRSKLSKDPKEKRDHMVLLEFVTAAGIKHGRDERYKTGGSGGSGGSGGSGGSGGSGGSGGTASNL
LTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAEPEDVR
DYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERT
DFDQVRSLMENSDRCQDIRNLAFLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAG
VEKALSLGVTKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAK
DDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGDG
GSHHHHHH (SEQ ID NO:33)

[00141] (–30)GFP-Cre-6xHis (-30 GFP double-underlined; 6xHis tag underlined):

MGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLVTTLTY
GVQCFSDYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDG
NILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLS
TESALSKDPNEDRDHMVLLEFVTAAGIDHGMDELYKTGGSGGSGGSGGSGGSGGSGGSGGSGGTASNL
LTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAEPEDVR
DYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERT
DFDQVRSLMENSDRCQDIRNLAFLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAG
VEKALSLGVTKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAK
DDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGDG
GSHHHHHH (SEQ ID NO:34)

[00142] (+36)GFP-PPARγ-TALE-2 (+36 GFP double-underlined; 6xHis tag underlined):

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLTY
GVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKG
NILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLS
TRSKLSKDPKEKRDHMVLLEFVTAAGIKHGRDERYKTGGSGGSGGSGGSGGSGGSGGSGGSGGTAPKK
KRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAV
KYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVH
AWRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETV
QRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETV
QRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETV
QRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETV
QRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETV
QRLLPVLCQAHGLTPEQVVAIASNIGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAV
KKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSGGGSGRADA
LDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDML<u>HHHHHH</u> (SEQ ID NO:36)

[00143] (+36)GFP-PRDM16_TALE-3(+36 GFP double-underlined; 6xHis tag underlined):

MGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPWPTLVTTLTY
GVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVNRIKLKGRDFKEKG
NILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLS
TRSKLSKDPKEKRDHMVLLEFVTAAGIKHGRDERYKTGGSGGSGGSGGSGGSGGSGGSGGSGGTAPKK
KRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAV
KYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVH
AWRNALTGAPLNLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETV
QRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETV
QRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETV
QRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETV
QRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETV
QRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETV
QRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETV
QRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETV
QRLLPVLCQAHGLTPEQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAV
KKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSGGGSGRADA

LDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDML<u>HHHHHH</u> (SEQ ID NO:37)

[00144] (-30)GFP-PPARγ-TALE-2 (-30 GFP double-underlined; 6xHis tag underlined):

<u>MGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLVTTLTY
GVQCFSDYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDG
NILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLS
TESALSKDPNEDRDHMVLLEFVTAAGIDHGMDELYK</u>APKKKRKVGIHRGVPMVDLRTLGYSQQQQEKI
KPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGAR
ALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNIGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGK
QALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGK
QALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGK
QALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGK
QALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGR
PALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAD
HAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSGGGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALD
DFDLDMLGSDALDDFDLDML<u>HHHHHH</u> (SEQ ID NO:38)

[00145] (-30)GFP-PRDM16_TALE-3(-30 GFP double-underlined; 6xHis tag underlined):

<u>MGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLVTTLTY
GVQCFSDYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDG
NILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLS
TESALSKDPNEDRDHMVLLEFVTAAGIDHGMDELYK</u>APKKKRKVGIHRGVPMVDLRTLGYSQQQQEKI
KPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGAR
ALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNGGGK
QALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGK
QALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGK
QALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGK
QALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGK
QALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGK
QALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGK
QALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGR
PALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAD
HAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSGGGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALD
DFDLDMLGSDALDDFDLDML<u>HHHHHH</u> (SEQ ID NO:39)

[00146] (-30)GFP:

MGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVPWPTLVTTLTY
GVQCFSDYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDG
NILGHKLEYNFNSHDVYITADKQENGIKAEFEIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDDHYLS
TESALSKDPNEDRDHMVLLEFVTAAGIDHGMDELYK (SEQ ID NO:40)

[00147] N-terminal TALE domain:

APKKKRKVGIHRGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALG TVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAV EAVHAWRNALTGAPLNL (SEQ ID NO:41)

[00148] C-terminal TALE domain:

LESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHA QVVRVLGFFQCHSHPAQAFDDAMTQFGMSGGGS (SEQ ID NO:42)

[00149] VP64 activation domain:

GRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDML (SEQ ID NO:35)

Compositions of functional effector proteins and cationic polymers

[00150] Certain aspects of the disclosure relate to the use of cationic polymers for the delivery of effector proteins (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*), for example as opposed to delivering "naked" protein preparations. As with cationic lipids, aspects of the present disclosure relate to the recognition that anionic proteins or protein complexes (including those proteins associated with nucleic acids) can take advantage of electrostatics-driven encapsulation by and/or association with cationic polymers for delivery of functional effector proteins. While few proteins natively possess the density of negative charges found in the phosphate backbone of nucleic acids, translational fusion to, or non-covalent association with, an anionic carrier such as a negatively supercharged protein or a nucleic acid as described herein render the resulting effector protein or protein complex sufficiently anionic to drive efficient encapsulation/association of such protein cargoes by cationic polymers.

[00151] In some embodiments, association or fusion with an engineered supernegatively charged GFP is capable of driving efficient encapsulation/association and delivery of proteins into cultured mammalian cells by cationic polymers. In some embodiments, Cas9 protein associated with polyanionic guide RNAs (gRNA) can be efficiently delivered in functional form into mammalian cells using cationic polymers. Accordingly, in some embodiments, a composition comprising a Cas9 protein and a cationic polymer is provided, wherein the Cas9 protein is associated with a gRNA, and the composition is capable of delivering the Cas9 protein to the interior of a cell. In some embodiments, delivery of Cas9:gRNA complexes using cationic polymers results in at least a

2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 20-fold or 25-fold or higher on-target:off-target modification ratio as compared with with plasmid transfection of the Cas9 protein.

[00152] Accordingly, some aspects of the disclosure provide compositions comprising a Cas9 protein (*e.g.*, as described herein; see *e.g.*, *Cas9 effector proteins* below) and a cationic polymer capable of delivering the Cas9 protein to the interior of a cell. In some embodiments, the Cas9 protein is associated with a gRNA, which *e.g.*, provides anionic charge to the complex thereby allowing the Cas9:gRNA complex to be encapsulated and/or associated with the cationic polymers. In some embodiments, the Cas9 protein need not be associated with a gRNA for effective encapsulation by and/or association with a cationic lipid, but instead is associated with a negatively supercharged protein, as described herein. In some embodiments where a Cas9 protein is associated with a negatively supercharged protein, the Cas9 protein is also associated with a gRNA. In some embodiments, the Cas9 protein is a wild type Cas9 protein, a fragment of a wild type Cas9 protein, or a variant of a wild type Cas9 protein. In some embodiments, the Cas9 protein comprises a dCas9 domain (*e.g.*, as described herein). In some embodiments, the Cas9 protein is a fusion protein comprising a dCas9 domain (*e.g.*, as described herein). In some embodiments, the Cas9 protein is a Cas9 nickase.

[00153] In other embodiments, compositions comprising an effector protein (*e.g.*, other than a Cas9 protein) and a cationic polymer are provided which are capable of delivering the effector protein to the interior of a cell (*e.g.*, to the nucleus of the cell). The effector protein is either naturally negatively charged, is modified to have a net overall negative charge, or is associated with a negatively supercharged protein, as described herein. In some embodiments, the effector protein is any effector protein described herein. In some embodiments, the effector protein is a recombinase, *e.g.*, any recombinase described herein. In some embodiments, the recombinase is Cre recombinase. In some embodiments, the Cre recombinase comprises the amino acid sequence of SEQ ID NO:32 (*e.g.*, with or without the 6xHis tag). In some embodiments, the Cre recombinase comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO:32 (*e.g.*, with or without the 6xHis tag). In some embodiments, the Cre recombinase is fused to a supercharged protein (*e.g.*, +36 GFP or -30GFP). In some embodiments, the Cre recombinase fused to a supercharged protein comprises the amino acid sequence of SEQ ID NO:33 (*e.g.*, with or without the 6xHis tag) or SEQ ID NO:34 (*e.g.*, with or without the 6xHis tag), or comprises an amino acid sequence

that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO:33 or SEQ ID NO:34 (*e.g.*, with or without the 6xHis tag). In some embodiments, the effector protein is a TALE protein, (*e.g.*, as described herein including those provided in the Examples). In some embodiments, the TALE protein comprises one or more of a VP64 transcriptional activator domain (*e.g.*, SEQ ID NO:35). In some embodiments, the TALE protein with a VP64 transcriptional activator domain further comprises an amino acid sequence selected from the group consisting of SEQ ID NO:36-39 (*e.g.*, with or without the 6xHis tag). In some embodiments, the TALE protein with a VP64 transcriptional activator domain comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:36-39 (*e.g.*, with or without the 6xHis tag). In some embodiments, the TALE effector protein comprises a (-30)GFP domain (*e.g.*, SEQ ID NO:21 or SEQ ID NO:40), a N-terminal region of a TALE domain (*e.g.*, SEQ ID NO:41), a variable repeat domain (*e.g.*, an 18.5mer repeat domain as provided in Maeder et al., "Robust, synergistic regulation of human gene expression using TALE activators." *Nat. Methods.* 2013; **10,** 243–245), a C-terminal TALE domain (*e.g.*, SEQ ID NO:42), a VP64 activation domain (*e.g.*, SEQ ID NO:35), and optionally one or more linkers (*e.g.*, GGS(9), SEQ ID NO: 252) between any domain and optionally a sequence tag (*e.g.*, 6xHis. SEQ ID NO: 253).

[00154] In some embodiments, the compositions comprising a Cas9 protein or other effector proteins described herein and a cationic polymer exhibit low toxicity when administered to a population of cells (*e.g.*, *in vitro* or *in vivo*). In some embodiments, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cells in a population are viable following administration of an inventive composition comprising a Cas9 protein or other effector protein and cationic polymers. Methods for assessing the toxicity of a composition when administered to a population of cells are well known in the art and include those described in the Examples.

Cas9 effector proteins

[00155] In some embodiments, effector proteins comprising a RNA-programmable protein (or fragment or variant thereof) is delivered to a target cell by a system or method provided herein. In some embodiments, an RNA-guided or RNA-programmable nuclease is delivered to a target cell by a system or method provided herein. In some embodiments, the RNA-programmable protein is a Cas9 nuclease, a Cas9 variant, or a fusion of a Cas9 protein,

which is delivered to a target cell by a system or method provided herein.

[00156] In some embodiments, the RNA-programmable nuclease is a (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csn1) from *Streptococcus pyogenes* (see, *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference. Because RNA-programmable nucleases (*e.g.*, Cas9) use RNA:DNA hybridization to determine target DNA cleavage sites, these proteins are able to cleave, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (*e.g.*, to modify a genome) are known in the art (see *e.g.*, Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013); Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013); Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31**, 227-229 (2013); Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013); Dicarlo, J.E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. *et al.* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

[00157] A Cas9 nuclease may also be referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The

tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand that is not complementary to the crRNA is first cut endonucleolytically, then trimmed 3′-5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNA. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference).

[00158] Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, proteins comprising Cas9 proteins or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain, an N-terminal

domain or a C-terminal domain, *etc.*), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to the corresponding fragment of wild type Cas9. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, SEQ ID NO:43 (nucleotide); SEQ ID NO:44 (amino acid)). In some embodiments, a Cas9 protein has an inactive (*e.g.*, an inactivated) DNA cleavage domain. A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease "dead" Cas9). In some embodiments, dCas9 corresponds to, or comprises in part or in whole, the amino acid set forth as SEQ ID NO:45, below. In some embodiments, variants of dCas9 (*e.g.*, variants of SEQ ID NO:45) are provided. For example, in some embodiments, variants having mutations other than D10A and H840A are provided, which result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domain of Cas9 (*e.g.*, substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of dCas9 (*e.g.*, variants of SEQ ID NO:45) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO:45. In some embodiments, variants of dCas9 (*e.g.*, variants of SEQ ID NO:45) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO:45, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more. In some embodiments, Cas9 "nickases" are provided which comprise a mutation which inactivates a single nuclease domain in Cas9. Such nickases induce a single strand break in a target nucleic acid as opposed to a double strand break.

[00159] Cas9

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTAT
AAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCT
CTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGG
AAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGA
CTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAA
GTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGAT
TTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAAT
CCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCT

ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTC
ATTGCTCAGCTCCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCT
AATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTA
GATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATT
TTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTAC
GATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATC
TTTTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTT
ATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGC
AAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGA
CAAGAAGACTTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTAT
TATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCA
TGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAA
AATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACA
AAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGAT
TTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTT
GATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATT
ATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTA
TTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAG
CTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCT
GGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGAT
AGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCT
AACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTA
ATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAAT
TCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTT
GAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAA
GAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCA
ATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTC
AAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACG
AAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAA
ATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGA
GAGGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGT
GAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATAT
CCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAG
CAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACA
CTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAA
GGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAG
ACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGG
GATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAA
GGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAA
AATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATAT
AGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTG
GCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGAT
AACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTT
TCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCA
ATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATAT
TTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCC
ATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:43)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL

TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO:44)
(single underline: HNH domain; double underline: RuvC domain)

[00160] dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO:45)
(single underline: HNH domain; double underline: RuvC domain)

[00161] Cas9 nickase (D10A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD
LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF
IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY
YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKV
MGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV
REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI
TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD
WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD (SEQ ID NO:46)

[00162] In some embodiments, fusion proteins comprising a Cas9 protein are provided

for use in any of the compositions and methods described herein. In some embodiments, the fusion protein comprises a dCas9 protein (*e.g.*, as described herein). In some embodiments, the fusion protein comprises a linker (*e.g.*, as described herein) between dCas9 and one or more domains (*e.g.*, enzymatic domains). In some embodiments, the fusion protein comprises dCas9 and a transcriptional activator domain, a transcriptional repressor domain, a recombinase domain, a gene editing domain (*e.g.*, a deaminase doman), or an epigenetic modifier domain.

[00163] In some embodiments, the general architecture of exemplary fusion proteins provided herein comprises the structure:

$[NH_2]$-[enzymatic domain]-[dCas9]-[COOH] or

$[NH_2]$-[dCas9]-[enzymatic domain]-[COOH];

wherein NH_2 is the N-terminus of the fusion protein, COOH is the C-terminus of the fusion protein, and the enzymatic domain comprises a nuclease domain (*e.g.*, FokI), a recombinase catalytic domain (*e.g.*, Hin, Gin, or Tn3 recombinase domains), a nucleic acid-editing domain (*e.g.*, a deaminase domain), a transcriptional activator domain (*e.g.*, VP64, p65), a transcriptional repressor domain (*e.g.*, KRAB, SID), or an epigenetic modifier (*e.g.*, LSD1 histone demethylase, TET1 hydoxylase).

[00164] Additional features may be present, for example, one or more linker sequences between certain domains. Other exemplary features that may be present are localization sequences, such as nuclear localization sequences (NLS; *e.g.*, MAPKKKRKVGIHRGVP (SEQ ID NO:47)); cytoplasmic localization sequences; export sequences, such as nuclear export sequences; or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable localization signal sequences and sequences of protein tags are provided herein and are known in the art, and include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags (*e.g.*, 3xFLAG TAG: MDYKDHDGDYKDHDIDYKDDDDK (SEQ ID NO:48)), hemagglutinin (HA) tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST) tags, green fluorescent protein (GFP) tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags , biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art.

[00165] In some embodiments, the enzymatic domain comprises a nuclease or a catalytic domain thereof. For example, in some embodiments, the general architecture of

exemplary ligand-dependent dCas9 fusion proteins with a nuclease domain comprises the structure:

$[NH_2]$-[NLS]-[dCas9]-[nuclease]-[COOH],

$[NH_2]$-[NLS]-[nuclease]-[dCas9]-[COOH],

$[NH_2]$-[dCas9]-[nuclease]-[COOH], or

$[NH_2]$-[nuclease]-[dCas9]-[COOH];

wherein NLS is a nuclear localization signal, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the nuclease domain. In some embodiments, a linker is inserted between the NLS and the nuclease and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the nuclease and/or the dCas9 domain. In some embodiments, the NLS is located between the nuclease and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some aspects, the nuclease domain is a nuclease requiring dimerization (*e.g.*, the coming together of two monomers of the nuclease) in order to cleave a target nucleic acid (*e.g.*, DNA). In some embodiments, the nuclease domain is a monomer of the FokI DNA cleavage domain. The FokI DNA cleavage domain is known, and in some aspects corresponds to amino acids 388–583 of FokI (NCBI accession number J04623). In some embodiments, the FokI DNA cleavage domain corresponds to amino acids 300-583, 320-583, 340-583, or 360-583 of FokI. See also Wah *et al.*, "Structure of FokI has implications for DNA cleavage" *Proc. Natl. Acad. Sci. USA.* 1998; 1;95(18):10564-9; Li *et al.*, "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain" *Nucleic Acids Res.* 2011; 39(1):359-72; Kim *et al.*, "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain" *Proc. Natl Acad. Sci. USA.* 1996; 93:1156–1160; the entire contents of each are herein incorporated by reference). In some embodiments, the FokI DNA cleavage domain corresponds to, or comprises in part or whole, the amino acid sequence set forth as SEQ ID NO:49. In some embodiments, the FokI DNA cleavage domain is a variant of FokI (*e.g.*, a variant of SEQ ID NO:49), as described herein. Other exemplary compositions and methods of using dCas9-nuclease fusion proteins can be found in U.S. patent application U.S.S.N 14/320,498; titled "Cas9-FokI fusion Proteins and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

[00166] FokI nuclease domain:

GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTV
GSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQ
LTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:49)

[00167] fCas9 (*e.g.*, dCas9-NLS-GGS3linker-FokI):

ATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCC
CTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC
AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAG
GTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAAT
CCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTG
ATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCA
AATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTC
GACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTT
ATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGA
AAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGG
CAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTAC
TATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
TGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAG
AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACG
AAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
CTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTC
GATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
ATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTC
TTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGT
GGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGAC
TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCG
AATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
ATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAA
AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCT
GTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGAT
TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTC
GTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTA
ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGC
CAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
CGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
AGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAA
TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATC
ACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
AAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
TGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAG
AAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
AAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAG
TATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAA
CTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAA
GATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAA
TTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAG
TATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAA

TCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGG
AAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC
AAGGCTGCAGGATCAGGTGGAAGTGGCGGCAGCGGAGGTTCTGGATCCCAACTAGTCAAAAGTGAACTGGAGGAG
AAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAATTGAAATTGCCAGAAAT
TCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATATAGAGGTAAACAT
TTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTGATCGTGGAT
ACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTCGAAGAAAAT
CAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTT
TTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAAT
GGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAA
GTCAGACGGAAATTTAATAACGGCGAGATAAACTTT (SEQ ID NO:50)

[00168] fCas9 (*e.g.*, NLS- dCas9-GGS3linker–FokI):

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCC
CCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGATAAAAAGTATTCTATTGGTTTAGCTATC
GGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGG
AACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCG
ACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTT
AGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAG
AAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTAT
CACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATG
ATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATC
CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATT
CTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGG
TTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGAT
GCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTAT
GCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAG
ATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAG
GCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGT
TATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACG
GAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCA
CATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAAT
CGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGG
TTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCG
TCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCAC
AGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAA
CCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTT
AAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGA
TTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAG
AATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTA
AAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGA
TTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGAC
GGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCA
CAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGC
ATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATC
GAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAG
GGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTT
TACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTAC
GACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGAT
AAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTC
CTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTT
GACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGAT
TCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAA
TTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGAC
GCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGT
GATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATAC
TTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCT
TTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTT
TTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTT

CCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGC
CCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAA
GAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGT
TACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAA
CGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTG
TATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAG
CACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAAT
CTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCAT
TTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATAC
ACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGAT
TTGTCACAGCTTGGGGGTGACTCAGGTGGAAGTGGCGGCAGCGGAGGTTCTGGATCCCAACTAGTCAAAAGTGAA
CTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAATTGAAATT
GCCAGAAATTCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATATAGA
GGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTG
ATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTC
GAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAA
TTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTAAATCATATCACT
AATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACC
TTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTT (SEQ ID NO:51)

[00169] fCas9 (*e.g.*, FokI-GGS3linker-dCas9-NLS):

ATGGGATCCCAACTAGTCAAAAGTGAACTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCT
CATGAATATATTGAATTAATTGAAATTGCCAGAAATTCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAA
TTTTTTATGAAAGTTTATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACT
GTCGGATCTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGC
CAAGCAGATGAAATGCAACGATATGTCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGG
AAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCT
CAGCTTACACGATTAAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGA
GAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTGGC
GGTAGTGGGGGATCTGGGGGAAGTATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGGA
TGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCG
ATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACC
GCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAA
GTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCC
ATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTA
GTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCAC
TTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTAT
AATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCT
AAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATA
GCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGT
AAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCT
GCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTA
TCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAA
CTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCG
AGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAA
CTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGC
GAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAA
ATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGA
AAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATC
GAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTAT
TTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGA
GAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGAC
TACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGT
ACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAA
GATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTG
TTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATC
AACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAAC
TTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGG
GACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAA

GTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAAT
CAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGC
AGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAA
AATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTA
CCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGT
GACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATA
ACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATT
AAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAA
TACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGA
AAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC
GTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGAC
GTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATT
ATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGG
GAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTC
AACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGAT
AAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCT
GTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACG
ATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAG
GATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCC
GGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTAC
GAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGAC
GAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGC
GCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAAC
CTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTG
CTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGT
GACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGAC
ATCGATTACAAGGATGACGATGACAAGGCTGCAGGA (SEQ ID NO:52)

[00170] fCas9 (*e.g.*, NLS -FokI-GGS3linker-dCas9):

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCC
CCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTGGAGGTTCTATGGGATCCCAACTAGTCAAAAGT
GAACTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAATTGAA
ATTGCCAGAAATTCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATAT
AGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGT
GTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATAT
GTCGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACG
GAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTAAATCATATC
ACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTA
ACCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTGGCGGTAGTGGGGGATCTGGGGGAAGT
ATGGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCC
CTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC
AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAG
GTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGAC
CTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAAT
CCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTG
ATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCA
AATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTC
GACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATC
CTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTAC
GATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTT
ATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGA
AAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGG
CAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTAC
TATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCA
TGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAG

AATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACG
AAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
CTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTC
GATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATA
ATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTC
TTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAG
TTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGT
GGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGAC
TCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCG
AATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTC
ATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAA
AACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCT
GTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGAT
TCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTC
GTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTA
ACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGC
CAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATT
CGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTT
AGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAA
TACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATC
ACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGAT
AAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTG
CAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGAC
TGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAG
AAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAA
AAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAG
TATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAA
CTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAA
GATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAA
TTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAG
TATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAA
TCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGAC (SEQ ID NO:53)

[00171] fCas9:

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCC
CCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTGGAGGTTCTGGATCCCAACTAGTCAAAAGTGAA
CTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAATTGAAATT
GCCAGAAATTCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATATAGA
GGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTG
ATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTC
GAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAA
TTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTAAATCATATCACT
AATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACC
TTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTAGCGGCAGCGAGACTCCCGGGACCTCAGAG
TCCGCCACACCCGAAAGTGATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTAATTCCGTTGGATGGGCTGTC
ATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAG
AATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGA
AGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGAT
TCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGA
AACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCA
ACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATT
GAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTG
TTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGA
CGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCA
CTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACG
TACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAAC

CTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCA
ATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAG
AAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAG
GAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGC
GAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCAT
GCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACC
TTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAA
GAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATG
ACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTG
TACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAG
AAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAG
AAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCAT
GACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG
TTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGAT
AAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATA
AGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAG
CTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTG
CACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGAT
GAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACT
CAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATC
TTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGG
GACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCC
TTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTT
CCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGA
AAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAG
CTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAG
AACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTT
CAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACC
GCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAG
ATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTC
TTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGT
GAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTA
AAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATC
GCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTA
GTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAG
CGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATA
ATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTT
CAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTG
AAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATA
GAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAAC
AAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCT
CCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCG
ACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGAC (SEQ ID NO:54)

[00172] In some embodiments, the enzymatic domain comprises a recombinase or catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary ligand-dependent dCas9 fusion proteins with a recombinase domain comprises the structure:

$[NH_2]$-[NLS]-[dCas9]-[recombinase]-[COOH],

$[NH_2]$-[NLS]-[recombinase]-[dCas9]-[COOH],

$[NH_2]$-[dCas9]-[recombinase]-[COOH], or

$[NH_2]$-[recombinase]-[dCas9]-[COOH];

wherein NLS is a nuclear localization signal, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the recombinase domain. In some embodiments, a linker is inserted between the NLS and the recombinase and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the recombinase domain and/or the dCas9 domain. In some embodiments, the NLS is located between the recombinase domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. By "catalytic domain of a recombinase," it is meant that a fusion protein includes a domain comprising an amino acid sequence of (*e.g.*, derived from) a recombinase, such that the domain is sufficient to induce recombination when contacted with a target nucleic acid (either alone or with additional factors including other recombinase catalytic domains which may or may not form part of the fusion protein). In some embodiments, a catalytic domain of a recombinase does not include the DNA binding domain of the recombinase. In some embodiments, the catalytic domain of a recombinase includes part or all of a recombinase, *e.g.*, the catalytic domain may include a recombinase domain and a DNA binding domain, or parts thereof, or the catalytic domain may include a recombinase domain and a DNA binding domain that is mutated or truncated to abolish DNA binding activity. Recombinases and catalytic domains of recombinases are known to those of skill in the art, and include, for example, those described herein. In some embodiments, the catalytic domain is derived from any recombinase. In some embodiments, the recombinase catalytic domain is a catalytic domain of aTn3 resolvase, a Hin recombinase, or a Gin recombinase. In some embodiments, the catalytic domain comprises a Tn3 resolvase (*e.g.*, Stark Tn3 recombinase) that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:55, as provided below. In some embodiments, a Tn3 catalytic domain is encoded by a variant of SEQ ID NO:55. In some embodiments, a Tn3 catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:56. In some embodiments, the catalytic domain comprises a Hin recombinase that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:57, as provided below. In some embodiments, a Hin catalytic domain is encoded by a variant of SEQ ID NO:57. In some embodiments, a Hin catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:58. In some embodiments, the catalytic domain comprises a Gin recombinase (*e.g.*, Gin beta recombinase) that is encoded by a nucleotide sequence comprising, in part or in whole, SEQ ID NO:59, as provided below. In some embodiments, a Gin catalytic domain is encoded by a variant of SEQ ID NO:59. In some embodiments, a

Gin catalytic domain is encoded by a polynucleotide (or a variant thereof) that encodes the polypeptide corresponding to SEQ ID NO:60. Other exemplary compositions and methods of using dCas9-recombinase fusion proteins can be found in U.S. patent application U.S.S.N 14/320,467; titled "Cas9 Variants and Uses Thereof," filed June 30, 2014; the entire contents of which are incorporated herein by reference.

[00173] Stark Tn3 recombinase (nucleotide: SEQ ID NO:55; amino acid: SEQ ID NO:56):

ATGGCCCTGTTTGGCTACGCACGCGTGTCTACCAGTCAACAGTCACTCGATTTGCAAGTGAGGGCTCTTAAAGAT
GCCGGAGTGAAGGCAAACAGAATTTTTACTGATAAGGCCAGCGGAAGCAGCACAGACAGAGAGGGGCTGGATCTC
CTGAGAATGAAGGTAAAGGAGGGTGATGTGATCTTGGTCAAAAAATTGGATCGACTGGGGAGAGACACAGCTGAT
ATGCTTCAGCTTATTAAAGAGTTTGACGCTCAGGGTGTTGCCGTGAGGTTTATCGATGACGGCATCTCAACCGAC
TCCTACATTGGTCTTATGTTTGTGACAATTTTGTCCGCTGTGGCTCAGGCTGAGCGGAGAAGGATTCTCGAAAGG
ACGAATGAGGGACGGCAAGCAGCTAAGTTGAAAGGTATCAAATTTGGCAGACGAAGG (SEQ ID NO:55)

MALFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVKEGDVILVKKLDRLGRDTAD
MLQLIKEFDAQGVAVRFIDDGISTDSYIGLMFVTILSAVAQAERRRILERTNEGRQAAKLKGIKFGRRR (SEQ ID NO:56)

[00174] Hin Recombinase (nucleotide: SEQ ID NO:57; amino acid: SEQ ID NO:58):

ATGGCAACCATTGGCTACATAAGGGTGTCTACCATCGACCAAAATATCGACCTGCAGCGCAACGCTCTGACATCC
GCCAACTGCGATCGGATCTTCGAGGATAGGATCAGTGGCAAGATCGCCAACCGGCCCGGTCTGAAGCGGGCTCTG
AAGTACGTGAATAAGGGCGATACTCTGGTTGTGTGGAAGTTGGATCGCTTGGGTAGATCAGTGAAGAATCTCGTA
GCCCTGATAAGCGAGCTGCACGAGAGGGGTGCACATTTCCATTCTCTGACCGATTCCATCGATACGTCTAGCGCC
ATGGGCCGATTCTTCTTTTACGTCATGTCCGCCCTCGCTGAAATGGAGCGCGAACTTATTGTTGAACGGACTTTG
GCTGGACTGGCAGCGGCTAGAGCACAGGGCCGACTTGGA (SEQ ID NO:57)

MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPGLKRALKYVNKGDTLVVWKLDRLGRSVKNLV
ALISELHERGAHFHSLTDSIDTSSAMGRFFFYVMSALAEMERELIVERTLAGLAAARAQGRLG (SEQ ID NO:58)

[00175] Gin beta recombinase (nucleotide: SEQ ID NO:59; amino acid: SEQ ID NO:60):

ATGCTCATTGGCTATGTAAGGGTCAGCACCAATGACCAAAACACAGACTTGCAACGCAATGCTTTGGTTTGCGCC
GGATGTGAACAGATATTTGAAGATAAACTGAGCGGCACTCGGACAGACAGACCTGGGCTTAAGAGAGCACTGAAA
AGACTGCAGAAGGGGGACACCCTGGTCGTCTGGAAACTGGATCGCCTCGGACGCAGCATGAAACATCTGATTAGC
CTGGTTGGTGAGCTTAGGGAGAGAGGAATCAACTTCAGAAGCCTGACCGACTCCATCGACACCAGTAGCCCCATG
GGACGATTCTTCTTCTATGTGATGGGAGCACTTGCTGAGATGGAAAGAGAGCTTATTATCGAAAGAACTATGGCT
GGTATCGCTGCTGCCCGGAACAAAGGCAGACGGTTCGGCAGACCGCCGAAGAGCGGC (SEQ ID NO:59)

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLDRLGRSMKHLIS
LVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGIAAARNKGRRFGRPPKSG (SEQ ID NO:60)

[00176] In some embodiments, the enzymatic domain comprises a deaminase or a catalytic domain thereof. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a deaminase enzyme or domain comprises the structure:

$[NH_2]$-[NLS]-[Cas9]-[deaminase]-[COOH],

$[NH_2]$-[NLS]-[deaminase]-[Cas9]-[COOH],

$[NH_2]$-[Cas9]-[deaminase]-[COOH], or

$[NH_2]$-[deaminase]-[Cas9]-[COOH];

wherein NLS is a nuclear localization signal, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the deaminase domain. In some embodiments, a linker is inserted between the NLS and the deaminase and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the deaminase and/or the dCas9 domain. In some embodiments, the NLS is located between the deaminase domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. One exemplary suitable type of nucleic acid-editing enzymes and domains are cytosine deaminases, for example, of the apolipoprotein B mRNA-editing complex (APOBEC) family of cytosine deaminase enzymes, including activation-induced cytidine deaminase (AID) and apolipoprotein B editing complex 3 (APOBEC3) enzyme. Another exemplary suitable type of nucleic acid-editing enzyme and domain thereof suitable for use in the present invention include adenosine deaminases. For example, an ADAT family adenosine deaminase can be fused to a dCas9 domain. Some exemplary suitable nucleic-acid editing enzymes and domains, *e.g.*, deaminases and deaminase domains, that can be fused to dCas9 domains according to aspects of this disclosure are provided below. It will be understood that, in some embodiments, the active domain of the respective sequence can be used, *e.g.*, the domain without a localizing signal (nuclear localizing signal, without nuclear export signal, cytoplasmic localizing signal). Other exemplary compositions and methods of using dCas9-nuclease fusion proteins can be found in U.S. patent application U.S.S.N 14/325,815; titled "Fusions of Cas9 Domains and Nucleic Acid-Editing Domains," filed July 8, 2014; the entire contents of which are incorporated herein by reference.

[00177] Human AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRC
YRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNT

FVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO:61)
(underline: nuclear localization signal; double underline: nuclear export signal)

[00178] Mouse AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRC
YRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNT
FVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO:62)
(underline: nuclear localization signal; double underline: nuclear export signal)

[00179] Dog AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRC
YRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNT
FVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO:63)
(underline: nuclear localization signal; double underline: nuclear export signal)

[00180] Bovine AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRC
YRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWN
TFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO:64)
(underline: nuclear localization signal; double underline: nuclear export signal)

[00181] Mouse APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEIC
FLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEG
AQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETR
FCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSM
ELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDC
WTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO:65)
(underline: nucleic acid editing domain)

[00182] Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEIC
FLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHHNLSLDIFSSRLYNIRDPENQQNLCRLVQEG
AQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETR
FCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSM
ELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDC
WTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDLVNDFGNLQLGPPMS (SEQ ID NO:66)
(underline: nucleic acid editing domain)

[00183] Rhesus macaque APOBEC-3G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY***HPEMRFLRWFHKWRQLHH***
DQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEF
QDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTW
VPLNQHRGFLRNQAPNIHGFPKGR***HAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSC***AQEMAKFISNNEHVSLC
IFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI

(SEQ ID NO:67)
(bold italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00184] Chimpanzee APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKYHPEMRFFHWFS
KWRKLHRDQEYEVTWYISWSPCTKCTRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMK
IMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVE
RLHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSCAQEMAKFIS
NNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLEEHSQALSGRL
RAILQNQGN (SEQ ID NO:68)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)

[00185] Green monkey APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKDHPEMKFLHWFR
KWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQERGGPHATMK
IMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVE
RSHNDTWVLLNQHRGFLRNQAPDRHGFPKGRHAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSCAQKMAKFISN
NKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLR
AI (SEQ ID NO:69)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)

[00186] Human APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFFHWFS
KWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMK
IMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVE
RMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFIS
KNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRL
RAILQNQEN (SEQ ID NO:70)
(underline: nucleic acid editing domain; double underline: cytoplasmic localization signal)

[00187] Human APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCFLSWFC
GNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDE
EFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVV
KHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLT
IFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE
(SEQ ID NO:71)
(underline: nucleic acid editing domain)

[00188] Human APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEMCFLSWF
CGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVTIMDY
EEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDN
GTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRA
ILQNQGN (SEQ ID NO:72)
(underline: nucleic acid editing domain)

[00189] Human APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAERCFLSWF
CDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDY
EDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ (SEQ ID NO:73)
(underline: nucleic acid editing domain)

[00190] Human APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRF
LDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQV
SIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO:74)
(underline: nucleic acid editing domain)

[00191] Human APOBEC-3H:

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQ
VTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVD
HEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO:75)
(underline: nucleic acid editing domain)

[00192] Human APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRF
ENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRL
HKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACG
RNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEV
AEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFR
LLKRRLREILQ (SEQ ID NO:76)
(underline: nucleic acid editing domain)

[00193] Human APOBEC-1:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERD
FHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYY
HCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLA
TGLIHPSVAWR (SEQ ID NO:77)

[00194] Mouse APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKFTTERY
FRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTEQEYC
YCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWA
TGLK (SEQ ID NO:78)

[00195] Rat APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERY
FCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESG
YCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWA
TGLK (SEQ ID NO:79)

[00196] Human ADAT-2:

MEAKAAPKPAASGACSVSAEETEKWMEEAMHMAKEALENTEVPVGCLMVYNNEVVGKGRNEVNQTKNATRHAEMV
AIDQVLDWCRQSGKSPSEVFEHTVLYVTVEPCIMCAAALRLMKIPLVVYGCQNERFGGCGSVLNIASADLPNTGR
PFQCIPGYRAEEAVEMLKTFYKQENPNAPKSKVRKKECQKS (SEQ ID NO:80)

[00197] Mouse ADAT-2:

MEEKVESTTTPDGPCVVSVQETEKWMEEAMRMAKEALENIEVPVGCLMVYNNEVVGKGRNEVNQTKNATRHAEMV
AIDQVLDWCHQHGQSPSTVFEHTVLYVTVEPCIMCAAALRLMKIPLVVYGCQNERFGGCGSVLNIASADLPNTGR
PFQCIPGYRAEEAVELLKTFYKQENPNAPKSKVRKKDCQKS (SEQ ID NO:81)

[00198] Mouse ADAT-1:

MWTADEIAQLCYAHYNVRLPKQGKPEPNREWTLLAAVVKIQASANQACDIPEKEVQVTKEVVSMGTGTKCIGQSK
MRESGDILNDSHAEIIARRSFQRYLLHQLHLAAVLKEDSIFVPGTQRGLWRLRPDLSFVFFSSHTPCGDASIIPM
LEFEEQPCCPVIRSWANNSPVQETENLEDSKDKRNCEDPASPVAKKMRLGTPARSLSNCVAHHGTQESGPVKPDV
SSSDLTKEEPDAANGIASGSFRVVDVYRTGAKCVPGETGDLREPGAAYHQVGLLRVKPGRGDRTCSMSCSDKMAR
WNVLGCQGALLMHFLEKPIYLSAVVIGKCPYSQEAMRRALTGRCEETLVLPRGFGVQELEIQQSGLLFEQSRCAV
HRKRGDSPGRLVPCGAAISWSAVPQQPLDVTANGFPQGTTKKEIGSPRARSRISKVELFRSFQKLLSSIADDEQP
DSIRVTKKLDTYQEYKDAASAYQEAWGALRRIQPFASWIRNPPDYHQFK (SEQ ID NO:82) (underline: nucleic acid editing domain)

[00199] Human ADAT-1:

MWTADEIAQLCYEHYGIRLPKKGKPEPNHEWTLLAAVVKIQSPADKACDTPDKPVQVTKEVVSMGTGTKCIGQSK
MRKNGDILNDSHAEVIARRSFQRYLLHQLQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSSHTPCGDASIIPM
LEFEDQPCCPVFRNWAHNSSVEASSNLEAPGNERKCEDPDSPVTKKMRLEPGTAAREVTNGAAHHQSFGKQKSGP
ISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSGKPGAAFHQVGLLRVKPGRGDRTRSMSCS
DKMARWNVLGCQGALLMHLLEEPIYLSAVVIGKCPYSQEAMQRALIGRCQNVSALPKGFGVQELKILQSDLLFEQ
SRSAVQAKRADSPGRLVPCGAAISWSAVPEQPLDVTANGFPQGTTKKTIGSLQARSQISKVELFRSFQKLLSRIA
RDKWPHSLRVQKLDTYQEYKEAASSYQEAWSTLRKQVFGSWIRNPPDYHQFK (SEQ ID NO:83)
(underline: nucleic acid editing domain)

[00200] In some embodiments, the enzymatic domain comprises one or more of a transcriptional activator. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a transcriptional activator domain comprises the structure:

$[NH_2]$-[NLS]-[Cas9]-[(transcriptional activator)$_n$]-[COOH],

$[NH_2]$-[NLS]-[(transcriptional activator)$_n$]-[Cas9]- -[COOH],

$[NH_2]$-[Cas9]-[(transcriptional activator)$_n$]-[COOH], or

$[NH_2]$-[(transcriptional activator)$_n$]-[Cas9]-[COOH];

wherein NLS is a nuclear localization signal, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion proteins comprises one or more repeats of the transcriptional activator, for example wherein n = 1-10 (*e.g.*, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10). In some embodiments, n = 1-20. In some

embodiments, a linker is inserted between the dCas9 and the transcriptional activator domain. In some embodiments, a linker is inserted between the NLS and the transcriptional activator and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the transcriptional activator and/or the dCas9 domain. In some embodiments, the NLS is located between the transcriptional activator domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some embodiments, the transcriptional activator is selected from the group consisting of VP64, (SEQ ID NO:84 or SEQ ID NO:35), VP16 (SEQ ID NO:85), and p65 (SEQ ID NO:86). In some embodiments, a dCas9-VP64 fusion protein comprises the amino acid sequence of SEQ ID NO:87 (*e.g.*, with or without the 6xHis tag) or comprises an amino acid sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to the amino acid sequence of SEQ ID NO:87 (*e.g.*, with or without the 6xHis tag).

[00201] VP64

GSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLIN (SEQ ID NO:84)

[00202] VP16

APPTDVSLGDELHLDGEDVAMAHADALDDFDLDMLGDGDSPGPGFTPHDSAPYGALDMADFEFEQMFT
DALGIDEYGGEFPGIRR (SEQ ID NO:85)

[00203] p65:

PSGQISNQALALAPSSAPVLAQTMVPSSAMVPLAQPPAPAPVLTPGPPQSLSAPVPKSTQAGEGTLSE
ALLHLQFDADEDLGALLGNSTDPGVFTDLASVDNSEFQQLLNQGVSMSHSTAEPMLMEYPEAITRLVT
GSQRPPDPAPTPLGTSGLPNGLSGDEDFSSIADMDFSALLSQISSSGQ (SEQ ID NO:86)

[00204] dCas9-VP64-6xHis:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK

MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAGGGGSGRADALDDFDLDMLGSDALD
DFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLHHHHHH (SEQ ID NO:87)

[00205] In some embodiments, the enzymatic domain comprises one or more of a transcriptional repressor. For example, in some embodiments, the general architecture of exemplary dCas9 fusion proteins with a transcriptional repressor domain comprises the structure:

$[NH_2]$-[NLS]-[Cas9]-[(transcriptional repressor)$_n$]-[COOH],

$[NH_2]$-[NLS]-[(transcriptional repressor)$_n$]-[Cas9]- -[COOH],

$[NH_2]$-[Cas9]-[(transcriptional repressor)$_n$]-[COOH], or

$[NH_2]$-[(transcriptional repressor)$_n$]-[Cas9]-[COOH];

wherein NLS is a nuclear localization signal, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion proteins comprises one or more repeats of the transcriptional repressor, for example wherein n = 1-10 (*e.g.*, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10). In some embodiments, n = 1-20. In some embodiments, a linker is inserted between the dCas9 and the transcriptional repressor domain. In some embodiments, a linker is inserted between the NLS and the transcriptional repressor and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the transcriptional repressor and/or the dCas9 domain. In some embodiments, the NLS is located between the transcriptional repressor domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. In some embodiments, the transcriptional repressor is selected from the group consisting of the KRAB (Krüppel associated box) domain of Kox1, SID (mSin3 interaction domain), the CS (Chromo Shadow) domain of HP1α, or the WRPW domain of Hes1. These and other repressor domains are known in the art, and in some embodiments correspond to those described in Urrutia, KRAB-containing zinc-finger repressor proteins. *Genome Biol.* 2003;4(10):231; Gilbert *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* 2013; 154, 442-451; Konermann et al., Optical control of mammalian endogenous transcription and epigenetic states. *Nature.* 2013; 500, 472-476; and published U.S. patent application U.S.S.N. 14/105,017, published as U.S. 2014/0186958 A1, the entire contents of which are incorporated herein by reference. In some embodiments, the transcription repressor domain

comprises one or more repeats (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, or 10 repeats) of a KRAB domain. In some embodiments, the KRAB domain comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:88-91. In some embodiments, the transcriptional repressor domains comprises one or more repeats of a SID protein. In some embodiments, the SID protein comprises an amino acid sequence set forth as SEQ ID NO:80. In some embodiments, the repressor domain comprises 2, 3, 4, 5, 6, 7, 8, 9, or 10 repeats of a SID protein (*e.g.*, SEQ ID NO:92). In some embodiments, the repressor domain comprises four repeats of SID (*e.g.*, SID4x; SEQ ID NO:93).

[00206] KRAB (human; GenBank: AAD20972.1)

MNMFKEAVTFKDVAVAFTEEELGLLGPAQRKLYRDVMVENFRNLLSVGHPPFKQDVSPIERNEQLWIM TTATRRQGNLDTLPVKALLLYDLAQT (SEQ ID NO:88)

[00207] KRAB protein domain, partial (human; GenBank: CAB52478.1):

EQVSFKDVCVDFTQEEWYLLDPAQKILYRDVILENYSNLVSVGYCITKPEVIFKIEQGEEPWILEKGF PSQCHP (SEQ ID NO:89)

[00208] KRAB A domain, partial (human; GenBank: AAB03530.1):

EAVTFKDVAVVFTEEELGLLDPAQRKLYRDVMLENFRNLLSV (SEQ ID NO:90)

[00209] KRAB (mouse; C2H2 type domain containing protein; GenBank: CAM27971.1):

MDLVTYDDVHVNFTQDEWALLDPSQKSLYKGVMLETYKNLTAIGYIWEEHTIEDHFQTSRSHGSNKKT H (SEQ ID NO:91)

[00210] SID repressor domain:

GSGMNIQMLLEAADYLERREREAEHGYASMLP (SEQ ID NO:92)

[00211] SID4x repressor domain:

GSGMNIQMLLEAADYLERREREAEHGYASMLPGSGMNIQMLLEAADYLERREREAEHGYASMLPGSGM NIQMLLEAADYLERREREAEHGYASMLPGSGMNIQMLLEAADYLERREREAEHGYASMLPSR (SEQ ID NO:93)

[00212] In some embodiments, the enzymatic domain comprises an epigenetic modifier or a catalytic domain thereof. For example, in some embodiments, the general

architecture of exemplary dCas9 fusion proteins with an epigenetic modifier or domain comprises the structure:

$[NH_2]$-[NLS]-[Cas9]-[epigenetic modifier]-[COOH],

$[NH_2]$-[NLS]-[epigenetic modifier]-[Cas9]-[COOH],

$[NH_2]$-[Cas9]-[epigenetic modifier]-[COOH], or

$[NH_2]$-[epigenetic modifier]-[Cas9]-[COOH];

wherein NLS is a nuclear localization signal, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the dCas9 and the epigenetic modifier domain. In some embodiments, a linker is inserted between the NLS and the epigenetic modifier and/or dCas9 domain. In some embodiments, the NLS is located C-terminal of the epigenetic modifier and/or the dCas9 domain. In some embodiments, the NLS is located between the epigenetic modifier domain and the dCas9 domain. Additional features, such as sequence tags, may also be present. Epigenetic modifiers are well known in the art, and typically catalyze DNA methylation (and demethylation) or histone modifications (*e.g.*, histone methylation/demethylation, acetylation/deacetylation, ubiquitylation, phosphorylation, sumoylation, *etc.*). The presence of one more epigenetic modifications can affect the transcriptional activity of one or more genes, for example turning genes from an "on" state to an "off" state, and *vice versa*. Epigenetic modifiers include, but are not limited to, histone demethylase, histone methyltransferase, hydroxylase, histone deacetylase, and histone acetyltransferase. Exemplary epigenetic modifying proteins can be found in Konermann *et al.*, Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013; 500, 472-476; Mendenhall *et al.*, Locus-specific editing of histone modifications at endogenous enhancers. *Nat. Biotechnol.* 2013; 31, 1133-1136; and Maeder *et al.*, Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* 2013; 31, 1137-1142; the entire contents of each are incorporated herein by reference. In some embodiments, the epigenetic modifier domain is LSD1 (Lysine (K)-specific demethylase 1A) histone demethylase, which in some embodiments, comprises in whole or in part, an amino acid sequence set forth as SEQ ID NO:94 or SEQ ID NO:95. In some embodiments, the epigenetic modifier domain is TET1 hydroxylase catalytic domain, which in some embodiments, comprises an amino acid sequence set forth as SEQ ID NO:96. In some embodiments, the epigenetic modifier is a histone deacetylase (HDAC) effector domain. In some embodiments, the HDAC effector domain comprises in whole in in part, an amino acid sequence corresponding to any of the HDAC effector proteins provided in

Supplementary Table 2 of Konermann *et al*., Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013; 500, 472-476; SEQ ID NOs:97-108. In some embodiments, the epigenetic modifier is a histone methyltransferase (HMT) effector domain. In some embodiments, the HMT effector domain comprises in whole in in part, an amino acid sequence corresponding to any of the HDAC effector proteins provided in Supplementary Table 3 of Konermann *et al*., Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013; 500, 472-476; SEQ ID NOs:109-118.

[00213] LSD1, isoform a (human):

MLSGKKAAAAAAAAAAAATGTEAGPGTAGGSENGSEVAAQPAGLSGPAEVGPGAVGERTPRKKEPPRA
SPPGGLAEPPGSAGPQAGPTVVPGSATPMETGIAETPEGRRTSRRKRAKVEYREMDESLANLSEDEYY
SEEERNAKAEKEKKLPPPPPQAPPEEENESEPEEPSGQAGGLQDDSSGGYGDGQASGVEGAAFQSRLP
HDRMTSQEAACFPDIISGPQQTQKVFLFIRNRTLQLWLDNPKIQLTFEATLQQLEAPYNSDTVLVHRV
HSYLERHGLINFGIYKRIKPLPTKKTGKVIIIGSGVSGLAAARQLQSFGMDVTLLEARDRVGGRVATF
RKGNYVADLGAMVVTGLGGNPMAVVSKQVNMELAKIKQKCPLYEANGQADTVKVPKEKDEMVEQEFNR
LLEATSYLSHQLDFNVLNNKPVSLGQALEVVIQLQEKHVKDEQIEHWKKIVKTQEELKELLNKMVNLK
EKIKELHQQYKEASEVKPPRDITAEFLVKSKHRDLTALCKEYDELAETQGKLEEKLQELEANPPSDVY
LSSRDRQILDWHFANLEFANATPLSTLSLKHWDQDDDFEFTGSHLTVRNGYSCVPVALAEGLDIKLNT
AVRQVRYTASGCEVIAVNTRSTSQTFIYKCDAVLCTLPLGVLKQQPPAVQFVPPLPEWKTSAVQRMGF
GNLNKVVLCFDRVFWDPSVNLFGHVGSTTASRGELFLFWNLYKAPILLALVAGEAAGIMENISDDVIV
GRCLAILKGIFGSSAVPQPKETVVSRWRADPWARGSYSYVAAGSSGNDYDLMAQPITPGPSIPGAPQP
IPRLFFAGEHTIRNYPATVHGALLSGLREAGRIADQFLGAMYTLPRQATPGVPAQQSPSM (SEQ ID NO:94)

[00214] LSD1, isoform b (human):

MLSGKKAAAAAAAAAAAATGTEAGPGTAGGSENGSEVAAQPAGLSGPAEVGPGAVGERTPRKKEPPRA
SPPGGLAEPPGSAGPQAGPTVVPGSATPMETGIAETPEGRRTSRRKRAKVEYREMDESLANLSEDEYY
SEEERNAKAEKEKKLPPPPPQAPPEEENESEPEEPSGVEGAAFQSRLPHDRMTSQEAACFPDIISGPQ
QTQKVFLFIRNRTLQLWLDNPKIQLTFEATLQQLEAPYNSDTVLVHRVHSYLERHGLINFGIYKRIKP
LPTKKTGKVIIIGSGVSGLAAARQLQSFGMDVTLLEARDRVGGRVATFRKGNYVADLGAMVVTGLGGN
PMAVVSKQVNMELAKIKQKCPLYEANGQAVPKEKDEMVEQEFNRLLEATSYLSHQLDFNVLNNKPVSL
GQALEVVIQLQEKHVKDEQIEHWKKIVKTQEELKELLNKMVNLKEKIKELHQQYKEASEVKPPRDITA
EFLVKSKHRDLTALCKEYDELAETQGKLEEKLQELEANPPSDVYLSSRDRQILDWHFANLEFANATPL
STLSLKHWDQDDDFEFTGSHLTVRNGYSCVPVALAEGLDIKLNTAVRQVRYTASGCEVIAVNTRSTSQ
TFIYKCDAVLCTLPLGVLKQQPPAVQFVPPLPEWKTSAVQRMGFGNLNKVVLCFDRVFWDPSVNLFGH
VGSTTASRGELFLFWNLYKAPILLALVAGEAAGIMENISDDVIVGRCLAILKGIFGSSAVPQPKETVV
SRWRADPWARGSYSYVAAGSSGNDYDLMAQPITPGPSIPGAPQPIPRLFFAGEHTIRNYPATVHGALL
SGLREAGRIADQFLGAMYTLPRQATPGVPAQQSPSM (SEQ ID NO:95)

[00215] TET1 catalytic domain:

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQV
VRVLGFFQCHSHPAQAFDDAMTQFGMSGGGSLPTCSCLDRVIQKDKGPYYTHLGAGPSVAAVREIMEN
RYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVLIMVWD
GIPLPMADRLYTELTENLKSYNGHPTDRRCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFG
RSPSPRRFRIDPSSPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFS
GVTACLDFCAHPHRDIHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKEGMEA
KIKSGAIEVLAPRRKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRIKRKNNSTTTNNSKP

SSLPTLGSNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLK
NDATASCGFSERSSTPHCTMPSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPSTGVT
EPLTPHQPNHQPSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDSEH
IFLDANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRLSLVFYQHKNLNKPQHGFELNKIK
FEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTLTHDNVVTVSPYALTHVAGPYNHW
V (SEQ ID NO:96)

[00216] HDAC effector domains:

[00217] HDAC8 (*X. laevis*):

ASSPKKKRKVEASMSRVVKPKVASMEEMAAFHTDAYLQHLHKVSEEGDNDDPETLEYGLGYDCPITEG
IYDYAAAVGGATLTAAEQLIEGKTRIAVNWPGGWHHAKKDEASGFCYLNDAVLGILKLREKFDRVLYV
DMDLHHGDGVEDAFSFTSKVMTVSLHKFSPGFFPGTGDVSDIGLGKGRYYSINVPLQDGIQDDKYYQI
CEGVLKEVFTTFNPEAVVLQLGADTIAGDPMCSFNMTPEGIGKCLKYVLQWQLPTLILGGGGYHLPNT
ARCWTYLTALIVGRTLSSEIPDHEFFTEYGPDYVLEITPSCRPDRNDTQKVQEILQSIKGNLKRVVEF
(SEQ ID NO:97)

[00218] RPD3 (*S. cerevisiae*):

ASSPKKKRKVEASRRVAYFYDADVGNYAYGAGHPMKPHRIRMAHSLIMNYGLYKKMEIYRAKPATKQE
MCQFHTDEYIDFLSRVTPDNLEMFKRESVKFNVGDDCPVFDGLYEYCSISGGGSMEGAARLNRGKCDV
AVNYAGGLHHAKKSEASGFCYLNDIVLGIIELLRYHPRVLYIDIDVHHGDGVEEAFYTTDRVMTCSFH
KYGEFFPGTGELRDIGVGAGKNYAVNVPLRDGIDDATYRSVFEPVIKKIMEWYQPSAVVLQCGGDSLS
GDRLGCFNLSMEGHANCVNYVKSFGIPMMVVGGGGYTMRNVARTWCFETGLLNNVVLDKDLPYEF
(SEQ ID NO:98)

[00219] MesoLo4 (*M. loti*):

ASSPKKKRKVEASMPLQIVHHPDYDAGFATNHRFPMSKYPLLMEALRARGLASPDALNTTEPAPASWL
KLAHAADYVDQVISCSVPEKIEREIGFPVGPRVSLRAQLATGGTILAARLALRHGIACNTAGGSHHAR
RAQGAGFCTFNDVAVASLVLLDEGAAQNILVVDLDVHQGDGTADILSDEPGVFTFSMHGERNYPVRKI
ASDLDIALPDGTGDAAYLRRLATILPELSARARWDIVFYNAGVDVHAEDRLGRLALSNGGLRARDEMV
IGHFRALGIPVCGVIGGGYSTDVPALASRHAILFEVASTYAEF (SEQ ID NO:99)

[00220] HDAC11 (human):

ASSPKKKRKVEASMLHTTQLYQHVPETRWPIVYSPRYNITFMGLEKLHPFDAGKWGKVINFLKEEKLL
SDSMLVEAREASEEDLLVVHTRRYLNELKWSFAVATITEIPPVIFLPNFLVQRKVLRPLRTQTGGTIM
AGKLAVERGWAINVGGGFHHCSSDRGGGFCAYADITLAIKFLFERVEGISRATIIDLDAHQGNGHERD
FMDDKRVYIMDVYNRHIYPGDRFAKQAIRRKVELEWGTEDDEYLDKVERNIKKSLQEHLPDVVVYNAG
TDILEGDRLGGLSISPAGIVKRDELVFRMVRGRRVPILMVTSGGYQKRTARIIADSILNLFGLGLIGP
ESPSVSAQNSDTPLLPPAVPEF (SEQ ID NO:100)

[00221] HDT1 (*A. thaliana*):

ASSPKKKRKVEASMEFWGIEVKSGKPVTVTPEEGILIHVSQASLGECKNKKGEFVPLHVKVGNQNLVL
GTLSTENIPQLFCDLVFDKEFELSHTWGKGSVYFVGYKTPNIEPQGYSEEEEEEEEEVPAGNAAKAVA
KPKAKPAEVKPAVDDEEDESDSDGMDEDDSDGEDSEEEEPTPKKPASSKKRANETTPKAPVSAKKAKV
AVTPQKTDEKKKGGKAANQSEF (SEQ ID NO:101)

[00222] SIRT3 (human):

ASSPKKKRKVEASMVGAGISTPSGIPDFRSPGSGLYSNLQQYDLPYPEAIFELPFFFHNPKPFFTLAK
ELYPGNYKPNVTHYFLRLLHDKGLLLRLYTQNIDGLERVSGIPASKLVEAHGTFASATCTVCQRPFPG
EDIRADVMADRVPRCPVCTGVVKPDIVFFGEPLPQRFLLHVVDFPMADLLLILGTSLEVEPFASLTEA
VRSSVPRLLINRDLVGPLAWHPRSRDVAQLGDVVHGVESLVELLGWTEEMRDLVQRETGKLDGPDKEF
(SEQ ID NO:102)

[00223] HST2 (*S. cerevisiae*):

ASSPKKKRKVEASTEMSVRKIAAHMKSNPNAKVIFMVGAGISTSCGIPDFRSPGTGLYHNLARLKLPY
PEAVFDVDFFQSDPLPFYTLAKELYPGNFRPSKFHYLLKLFQDKDVLKRVYTQNIDTLERQAGVKDDL
IIEAHGSFAHCHCIGCGKVYPPQVFKSKLAEHPIKDFVKCDVCGELVKPAIVFFGEDLPDSFSETWLN
DSEWLREKITTSGKHPQQPLVIVVGTSLAVYPFASLPEEIPRKVKRVLCNLETVGDFKANKRPTDLIV
HQYSDEFAEQLVEELGWQEDFEKILTAQGGMGEF (SEQ ID NO:103)

[00224] CobB (*E. coli* (*K12*)):

ASSPKKKRKVEASMEKPRVLVLTGAGISAESGIRTFRAADGLWEEHRVEDVATPEGFDRDPELVQAFY
NARRRQLQQPEIQPNAAHLALAKLQDALGDRFLLVTQNIDNLHERAGNTNVIHMHGELLKVRCSQSGQ
VLDWTGDVTPEDKCHCCQFPAPLRPHVVWFGEMPLGMDEIYMALSMADIFIAIGTSGHVYPAAGFVHE
AKLHGAHTVELNLEPSQVGNEFAEKYYGPASQVVPEFVEKLLKGLKAGSIAEF (SEQ ID NO:104)

[00225] HST2 (*C. albicans*):

ASSPKKKRKVEASMPSLDDILKPVAEAVKNGKKVTFFNGAGISTGAGIPDFRSPDTGLYANLAKLNLP
FAEAVFDIDFFKEDPKPFYTLAEELYPGNFAPTKFHHFIKLLQDQGSLKRVYTQNIDTLERLAGVEDK
YIVEAHGSFASNHCVDCHKEMTTETLKTYMKDKKIPSCQHCEGYVKPDIVFFGEGLPVKFFDLWEDDC
EDVEVAIVAGTSLTVFPFASLPGEVNKKCLRVLVNKEKVGTFKHEPRKSDIIALHDCDIVAERLCTLL
GLDDKLNEVYEKEKIKYSKAETKEIKMHEIEDKLKEEAHLKEDKHTTKVDKKEKQNDANDKELEQLID
KAKAEF (SEQ ID NO:105)

[00226] SIRT5 (human):

ASSPKKKRKVEASSSSMADFRKFFAKAKHIVIISGAGVSAESGVPTFRGAGGYWRKWQAQDLATPLAF
AHNPSRVWEFYHYRREVMGSKEPNAGHRAIAECETRLGKQGRRVVVITQNIDELHRKAGTKNLLEIHG
SLFKTRCTSCGVVAENYKSPICPALSGKGAPEPGTQDASIPVEKLPRCEEAGCGGLLRPHVVWFGENL
DPAILEEVDRELAHCDLCLVVGTSSVVYPAAMFAPQVAARGVPVAEFNTETTPATNRFRFHFQGPCGT
TLPEALACHENETVSEF (SEQ ID NO:106)

[00227] Sir2A (*P. falciparum*):

ASSPKKKRKVEASMGNLMISFLKKDTQSITLEELAKIIKKCKHVVALTGSGTSAESNIPSFRGSSNSI
WSKYDPRIYGTIWGFWKYPEKIWEVIRDISSDYEIEINNGHVALSTLESLGYLKSVVTQNVDGLHEAS
GNTKVISLHGNVFEAVCCTCNKIVKLNKIMLQKTSHFMHQLPPECPCGGIFKPNIILFGEVVSSDLLK
EAEEEIAKCDLLLVIGTSSTVSTATNLCHFACKKKKKIVEINISKTYITNKMSDYHVCAKFSELTKVA
NILKGSSEKNKKIMEF (SEQ ID NO:107)

[00228] SIRT6 (human):

ASSPKKKRKVEASMSVNYAAGLSPYADKGKCGLPEIFDPPEELERKVWELARLVWQSSSVVFHTGAGI
STASGIPDFRGPHGVWTMEERGLAPKFDTTFESARPTQTHMALVQLERVGLLRFLVSQNVDGLHVRSG
FPRDKLAELHGNMFVEECAKCKTQYVRDTVVGTMGLKATGRLCTVAKARGLRACRGELRDTILDWEDS

LPDRDLALADEASRNADLSITLGTSLQIRPSGNLPLATKRRGGRLVIVNLQPTKHDRHADLRIHGYVD
EVMTRLMKHLGLEIPAWDGPRVLERALPPLEF (SEQ ID NO:108)

[00229] HMT effector domains:

[00230] NUE (*C. trachomatis*):

ASSPKKKRKVEASMTTNSTQDTLYLSLHGGIDSAIPYPVRRVEQLLQFSFLPELQFQNAAVKQRIQRL
CYREEKRLAVSSLAKWLGQLHKQRLRAPKNPPVAICWINSYVGYGVFARESIPAWSYIGEYTGILRRR
QALWLDENDYCFRYPVPRYSFRYFTIDSGMQGNVTRFINHSDNPNLEAIGAFENGIFHIIIRAIKDIL
PGEELCYHYGPLYWKHRKKREEFVPQEEEF (SEQ ID NO:109)

[00231] vSET (*P. bursaria chlorella virus*):

ASSPKKKRKVEASMFNDRVIVKKSPLGGYGVFARKSFEKGELVEECLCIVRHNDDWGTALEDYLFSRK
NMSAMALGFGAIFNHSKDPNARHELTAGLKRMRIFTIKPIAIGEEITISYGDDYWLSRPRLTQNEF
(SEQ ID NO:110)

[00232] SUV39H1 (human):

ASSPKKKRKVEASNLKCVRILKQFHKDLERELLRRHHRSKTPRHLDPSLANYLVQKAKQRRALRRWEQ
ELNAKRSHLGRITVENEVDLDGPPRAFVYINEYRVGEGITLNQVAVGCECQDCLWAPTGGCCPGASLH
KFAYNDQGQVRLRAGLPIYECNSRCRCGYDCPNRVVQKGIRYDLCIFRTDDGRGWGVRTLEKIRKNSF
VMEYVGEIITSEEAERRGQIYDRQGATYLFDLDYVEDVYTVDAAYYGNISHFVNHSCDPNLQVYNVFI
DNLDERLPRIAFFATRTIRAGEELTFDYNMQVDPVDMESTRMDSNFGLAGLPGSPKKRVRIECKCGTE
SCRKYLFEF (SEQ ID NO:111)

[00233] DIM5 (*N. crassa*):

ASSPKKKRKVEASMEKAFRPHFFNHGKPDANPKEKKNCHWCQIRSFATHAQLPISIVNREDDAFLNPN
FRFIDHSIIGKNVPVADQSFRVGCSCASDEECMYSTCQCLDEMAPDSDEEADPYTRKKRFAYYSQGAK
KGLLRDRVLQSQEPIYECHQGCACSKDCPNRVVERGRTVPLQIFRTKDRGWGVKCPVNIKRGQFVDRY
LGEIITSEEADRRRAESTIARRKDVYLFALDKFSDPDSLDPLLAGQPLEVDGEYMSGPTRFINHSCDP
NMAIFARVGDHADKHIHDLALFAIKDIPKGTELTFDYVNGLTGLESDAHDPSKISEMTKCLCGTAKCR
GYLWEF (SEQ ID NO:112)

[00234] KYP (*A. thaliana*):

ASSPKKKRKVEASDISGGLEFKGIPATNRVDDSPVSPTSGFTYIKSLIIEPNVIIPKSSTGCNCRGSC
TDSKKCACAKLNGGNFPYVDLNDGRLIESRDVVFECGPHCGCGPKCVNRTSQKRLRFNLEVFRSAKKG
WAVRSWEYIPAGSPVCEYIGVVRRTADVDTISDNEYIFEIDCQQTMQGLGGRQRRLRDVAVPMNNGVS
QSSEDENAPEFCIDAGSTGNFARFINHSCEPNLFVQCVLSSHQDIRLARVVLFAADNISPMQELTYDY
GYALDSVHEF (SEQ ID NO:113)

[00235] SUVR4 (*A. thaliana*):

ASSPKKKRKVEASQSAYLHVSLARISDEDCCANCKGNCLSADFPCTCARETSGEYAYTKEGLLKEKFL
DTCLKMKKEPDSFPKVYCKDCPLERDHDKGTYGKCDGHLIRKFIKECWRKCGCDMQCGNRVVQRGIRC
QLQVYFTQEGKGWGLRTLQDLPKGTFICEYIGEILTNTELYDRNVRSSSERHTYPVTLDADWGSEKDL
KDEEALCLDATICGNVARFINHRCEDANMIDIPIEIETPDRHYYHIAFFTLRDVKAMDELTWDYMIDF
NDKSHPVKAFRCCCGSESCRDRKIKGSQGKSIERRKIVSAKKQQGSKEVSKKRKEF (SEQ ID
NO:114)

[00236] Set4 (*C. elegans*):

ASSPKKKRKVEASMQLHEQIANISVTFNDIPRSDHSMTPTELCYFDDFATTLVVDSVLNFTTHKMSKK
RRYLYQDEYRTARTVMKTFREQRDWTNAIYGLLTLRSVSHFLSKLPPNKLFEFRDHIVRFLNMFILDS
GYTIQECKRYSQEGHQGAKLVSTGVWSRGDKIERLSGVVCLLSSEDEDSILAQEGSDFSVMYSTRKRC
STLWLGPGAYINHDCRPTCEFVSHGSTAHIRVLRDMVPGDEITCFYGSEFFGPNNIDCECCTCEKNMN
GAFSYLRGNENAEPIISEKKTKYELRSRSEF (SEQ ID NO:115)

[00237] Set1 (*C. elegans*):

ASSPKKKRKVEASMKVAAKKLATSRMRKDRAAAASPSSDIENSENPSSLASHSSSSGRMTPSKNTRSR
KGVSVKDVSNHKITEFFQVRRSNRKTSKQISDEAKHALRDTVLKGTNERLLEVYKDVVKGRGIRTKVN
FEKGDFVVEYRGVMMEYSEAKVIEEQYSNDEEIGSYMYFFEHNNKKWCIDATKESPWKGRLINHSVLR
PNLKTKVVEIDGSHHLILVARRQIAQGEELLYDYGDRSAETIAKNPWLVNTEF (SEQ ID NO:116)

[00238] SETD8 (human)

ASSPKKKRKVEASSCDSTNAAIAKQALKKPIKGKQAPRKKAQGKTQQNRKLTDFYPVRRSSRKSKAEL
QSEERKRIDELIESGKEEGMKIDLIDGKGRGVIATKQFSRGDFVVEYHGDLIEITDAKKREALYAQDP
STGCYMYYFQYLSKTYCVDATRETNRLGRLINHSKCGNCQTKLHDIDGVPHLILIASRDIAAGEELLY
DYGDRSKASIEAFPWLKHEF (SEQ ID NO:117)

[00239] TgSET8 (*T. gondii*):

ASSPKKKRKVEASASRRTGEFLRDAQAPSRWLKRSKTGQDDGAFCLETWLAGAGDDAAGGERGRDREG
AADKAKQREERRQKELEERFEEMKVEFEEKAQRMIARRAALTGEIYSDGKGSKKPRVPSLPENDDDAL
IEIIIDPEQGILKWPLSVMSIRQRTVIYQECLRRDLTACIHLTKVPGKGRAVFAADTILKDDFVVEYK
GELCSEREARE REQRYNRSKVPMGSFMFYFKNGSRMMAIDATDEKQDFGPARLINHSRRNPNMTPRAI
TLGDFNSEPRLIFVARRNIEKGEELLVDYGERDPDVIKEHPWLNSEF (SEQ ID NO:118)

[00240] Those of skill in the art will understand that any of the exemplary Cas9 proteins, including the exemplary Cas9 nucleases, variants, and fusions thereof, *e.g.*, described herein, can be delivered to cells using the instantly disclosed technology, and that the disclosure is not limited in this respect.

Nuclease effector proteins

[00241] TALE nucleases, or TALENs, are artificial nucleases comprising a transcriptional activator-like effector DNA binding domain associated with a DNA cleavage domain, for example, a FokI domain. A number of modular assembly schemes for generating engineered TALE constructs have been reported (Zhang, Feng; *et.al.* (February 2011). "Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription". *Nature Biotechnology* **29** (2): 149–53; Geiβler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. "Transcriptional Activators of Human Genes with Programmable DNA-Specificity". *PLoS ONE* **6** (5):

e19509; Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. *et al.* (2011). "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting". *Nucleic Acids Research*; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). "Assembly of custom TALE-type DNA binding domains by modular cloning". *Nucleic Acids Research*; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes". *Nucleic Acids Research.*; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). Bendahmane, Mohammed. ed. "Assembly of Designer TAL Effectors by Golden Gate Cloning". *PLoS ONE* **6** (5): e19722; the entire contents of each of which are incorporated herein by reference). Those of skill in the art will understand that TALE nucleases can be engineered to target virtually any genomic sequence with high specificity, and that such engineered nucleases can be used in embodiments of the present technology to manipulate the genome of a cell, *e.g.*, by delivering the respective TALEN via a method or strategy disclosed herein under circumstances suitable for the TALEN to bind and cleave its target sequence within the genome of the cell. In some embodiments, the delivered TALEN targets a gene or allele associated with a disease or disorder. In some embodiments, delivery of the TALEN to a subject confers a therapeutic benefit to the subject.

[00242] Zinc finger nucleases are a class of artificial nucleases that comprise a DNA cleavage domain and a zinc finger DNA binding domain. In some embodiments, the DNA cleavage domain is a non-specific DNA cleavage domain of a restriction endonuclease, for example, of FokI. In some embodiments, the DNA cleavage domain is a domain that only cleaves double-stranded DNA when dimerized with a second DNA cleavage domain of the same type. In some embodiments, the DNA cleavage domain is fused to the C-terminus of the zinc finger domain via a linker, for example, a peptide linker. In some embodiments, the zinc finger domain comprises between about 3 and about 6 zinc fingers and specifically recognizes and binds a target sequence of about 9-20 nucleotides in length. In some embodiments, a plurality of zinc finger nuclease molecules is delivered to a target cell by a system or method provided by this invention, with the zinc finger domain of one zinc finger nuclease molecule binding a target sequence in close proximity of the target sequence of a second zinc finger nuclease molecule. In some embodiments, the zinc finger domains of the zinc finger nuclease molecules binding target sequences in close proximity to each other are different. In some embodiments, a zinc finger nuclease molecule delivered to a cell by a system or method provided herein binds a target nucleic acid sequence in close proximity to

the target sequence of another zinc finger nuclease molecule, so that the DNA cleavage domains of the molecules dimerize and cleave a DNA molecule at a site between the two target sequences.

[00243] In some embodiments, the genome of the target cell is edited by a nuclease delivered to the cell via a strategy or method disclosed herein, *e.g.*, by a TALEN, or a zinc-finger nuclease, or a plurality or combination of such nucleases. In some embodiments, a single- or double-strand break is introduced at a specific site within the genome of a target cell by the nuclease, resulting in a disruption of the targeted genomic sequence. In some embodiments, the targeted genomic sequence is a nucleic acid sequence within the coding region of a gene. In some embodiments, the strand break introduced by the nuclease leads to a mutation within the target gene that impairs the expression of the encoded gene product. In some embodiments, a nucleic acid is co-delivered to the cell with the nuclease. In some embodiments, the nucleic acid comprises a sequence that is identical or homologous to a sequence adjacent to the nuclease target site. In some such embodiments, the strand break effected by the nuclease is repaired by the cellular DNA repair machinery to introduce all or part of the co-delivered nucleic acid into the cellular DNA at the break site, resulting in a targeted insertion of the co-delivered nucleic acid, or part thereof. In some embodiments, the insertion results in the disruption or repair of a pathogenic allele. In some embodiments, the insertion is detected by a suitable assay, *e.g.*, a DNA sequencing assay, a southern blot assay, or an assay for a reporter gene encoded by the co-delivered nucleic acid, *e.g.*, a fluorescent protein or resistance to an antibiotic. In some embodiments, the nucleic acid is co-delivered by association to a supercharged protein. In some embodiments, the supercharged protein is also associated to the functional effector protein, *e.g.*, the nuclease. In some embodiments, the delivery of a nuclease to a target cell results in a clinically or therapeutically beneficial disruption of the function of a gene.

[00244] In some embodiments, cells from a subject are obtained and a nuclease or other effector protein is delivered to the cells by a system or method provided herein *ex vivo*. In some embodiments, the treated cells are selected for those cells in which a desired nuclease-mediated genomic editing event has been effected. In some embodiments, treated cells carrying a desired genomic mutation or alteration are returned to the subject they were obtained from.

[00245] Methods for engineering, generation, and isolation of nucleases targeting specific sequences, *e.g.*, TALE, or zinc finger nucleases, and editing cellular genomes at specific target sequences, are well known in the art (see, *e.g.*, Mani *et al.*, *Biochemical and*

Biophysical Research Communications 335:447-457, 2005; Perez *et al.*, *Nature Biotechnology* 26:808-16, 2008; Kim *et al.*, *Genome Research*, 19:1279-88, 2009; Urnov *et al.*, *Nature* 435:646-51, 2005; Carroll *et al.*, *Gene Therapy* 15:1463-68, 2005; Lombardo *et al.*, *Nature Biotechnology* 25:1298-306, 2007; Kandavelou *et al.*, *Biochemical and Biophysical Research Communications* 388:56–61, 2009; and Hockemeyer *et al.*, *Nature Biotechnology* 27(9):851-59, 2009, as well as the reference recited in the respective section for each nuclease). The skilled artisan will be able to ascertain suitable methods for use in the context of the present disclosure based on the guidance provided herein.

TALE effector proteins

[00246] In some embodiments, effector proteins comprising a TALE domain are delivered to a target cell by a system or method provided herein. In some embodiments, a TALE effector, *e.g.*, an engineered TALE transcription factor comprising a TALE DNA binding domain and a heterologous transcriptional activator or repressor domain, is delivered to a cell by a system or method provided by aspects of this invention. In some embodiments, the TALE effector, *e.g.*, a transcription factor, is delivered to a cell in an amount sufficient to activate or inhibit transcription of a target gene of the transcription factor within the cell. In some embodiments, a transcription factor is delivered in an amount and over a time period sufficient to effect a change in the phenotype of a target cell, for example, a change in cellular function, or a change in developmental potential. Exemplary TALE transcription factors are described herein, and the skilled artisan will be able to identify additional suitable TALE transcription factors based on the guidance provided herein and the knowledge of such TALE transcription factors in the art.

[00247] In some embodiments, a target cell, for example, a somatic cell, is contacted with a TALE transcription factor, or a combination of such factors, associated with a supercharged protein provided herein. In some embodiments the target cell is a primary somatic cell and is contacted *in vitro* or *ex vivo* with a TALE transcription factor associated with a supercharged protein. In some embodiments, the TALE transcription factor is associated with a postiviely charged supercharged protein, *e.g.*, as described herein. In some embodiments, the TALE transcription factor is associated with a negatively charged supercharged proteins, *e.g.*, as described herein. In some embodiments, the TALE transcription factor is associated with a cationic lipid and/or cationic polymer, *e.g.*, as described herein. In some embodiments, the TALE transcription factor is associated with a negatively charged supercharged protein and a cationic lipid and/or cationic polymer, *e.g.*, as

described herein.

[00248] In some embodiments, a target cell is contacted, or repeatedly contacted, with a TALE transcription factor associated with a supercharged protein (and optionally a cationic lipid and/or cationic polymer) as provided herein, and a desired change in cellular phenotype or gene expression is detected. In some embodiments, a target cell is contacted repeatedly with a TALE transcription factor associated with a supercharged protein (and optionally a cationic lipid and/or cationic polymer) as provided herein until the formation of a desired cellular phenotype is detected. Methods for detecting cellular phenotypes and gene expression are well known to those in the art and include, for example, morphological analysis, and detection of marker gene expression by well-established methods such as immunohistochemistry, fluorescence activated cell sorting (FACS), or fluorescent microscopy. In some embodiments, a target cell is contacted with a TALE transcription factor associated with a supercharged protein as provided herein for a period of at least 3 hours, at least 6 hours, at least 12 hours, at least 1day, at least 2 days, at least 3 days, at least 4 days, at least 5 days, at least 6 days, at least 7 days, at least 10-12 days, at least 12-15 days, at least 15-20 days, at least 20-25 days, at least 25-30 days, at least 30-40 days, at least 40-50 days, at least 50-60 days, at least 60-70, or at least 70-100 days.

[0069] In some embodiments, a target cell is contacted with a TALE transcription factor associated with a supercharged protein (and optionally a cationic lipid and/or cationic polymer) as provided herein in an amount and for a time period effective to program the cell towards a different cell state. As will be apparent to those of skill in the art, the amount necessary to program or re-program a cell will dependent on various factors, for example, on the cell type and the treatment schedule. In general, delivery of a TALE transcription factor to a target somatic cell by a system or method provided herein will be at a concentration below a concentration at which significant toxicity can be observed. The critical concentration will depend, for example, on the specific TALE transcription factor, the supercharged protein it is associated with, the type of association, and the type of cell being treated.

[0070] A useful concentration of a functional effector protein associated with a supercharged protein (and optionally a cationic lipid and/or cationic polymer) for delivery to a specific cell type can be established by those of skill in the art by routine experimentation. In some embodiments a target cell is contacted *in vitro* or *ex vivo* with a functional effector protein associated with a supercharged protein (and optionally a cationic lipid and/or cationic polymer) at a concentration of about 1 pM to about 1 μM. In some embodiments, a target cell

is contacted *in vitro* or *ex vivo* with the functional effector protein associated to a supercharged protein at a concentration of about 1 pM, about 2.5 pM, about 5 pM, about 7.5 pM, about 10 pM, about 20 pM, about 25 pM, about 30 pM, about 40 pM, about 50 pM, about 60 pM, about 70 pM, about 75 pM, about 80 pM, about 90 pM, about 100 pM, about 200 pM, about 250 pM, about 300 pM, about 400 pM, about 500 pM, about 600 pM, about 700 pM, about 750 pM, about 800 pM, about 900 pM, about 1 nM, about 2nM, about 3nM, about 4nM, about 5nM, about 6nM, about 7 nM, about 8 nM, about 9nM, about 10nM, about 20nM, about 25 nM, about 30 nM, about 40 nM, about 50 nM, about 60 nm, about 70 nM, about 75 nM, about 80 nM, about 90 nM, about 100 nM, about 200 nM, about 250 nM, about 300 nM, about 400 nM, about 500 nM, about 600 nM, about 700 nM, about 750 nM, about 800 nM, about 900 nM, or about 1 μM. A useful time of exposure of the target cell to the functional effector protein, and, if necessary, incubation after administration in the absence of the functional effector protein, as well as a number of administration/incubation cycles useful to achieve a desired biological effect (*e.g.*, change in gene transcription, cleavage of a target site by a delivered nuclease, *etc.*), or a desired cellular phenotype can also be established by those of skill in the art by routine experimentation.

[0071] In some embodiments, the target cell for delivery of a functional effector protein by a system or method provided herein, is a primary cell obtained by a biopsy from a subject. In some embodiments, the subject is diagnosed as having a disease. In some embodiments the disease is a degenerative disease characterized by diminished function of a specific cell type, for example, a neural cell. In some embodiments, a cell treated with a functional effector protein according to the strategies or methods disclosed herein, or the progeny of such a cell, is used in a cell-replacement therapeutic approach. In some embodiments, the treated cells are administered to the subject from which the somatic cell was obtained in an autologous cell replacement therapeutic approach.

[0072] In some embodiments, a functional effector protein, *e.g.*, TALE transcription factor able to convert a cell from one differentiated state into another, is delivered to a target cell *in vitro* or *in vivo* by a system or method provided herein. Transcription factors that effect transdifferentiation are known in the art (see, *e.g.*, Zhou *et al.*, *Nature* 455:627-33, 2008). In some embodiments, a TALE transcription factor modulating the expression of PPARγ or PRDM16 are delivered to fibroblast cells by a system or method as provided by this invention. It is known in the art that expression these transcription factors is a pivotal step in the programming of fibroblasts towards a brown fat or white fat cell state. In some embodiments, a programmed brown fat cell is generated from a fibroblast obtained from a

subject in need of brown fat cells, and is administered to the subject, *e.g.*, used in a cell-replacement therapeutic approach involving the subject.

Formation of Complexes

[0073] The present invention provides complexes comprising supercharged proteins associated with one or more functional effector proteins to be delivered. In some embodiments, supercharged proteins are associated with one or more functional effector proteins to be delivered through non-covalent interactions. In some embodiments, supercharged proteins are associated with one or more functional effector proteins through electrostatic interactions. In certain embodiments, supercharged proteins have an overall net positive charge, and the functional effector proteins to be delivered have an overall net negative charge. In some embodiments, the complex further comprises a cationic lipid and/or cationic polymer. For example, in some embodiments, the supercharged protein of the complex is supernegatively charged, allowing for association with cationic lipids and/or polymers.

[0074] In certain embodiments, supercharged proteins are associated with one or more functional effector proteins to be delivered via covalent bond. For example, a supercharged protein may be fused to a functional effector protein to be delivered. Covalent attachment may be direct or indirect (*e.g.*, through a linker). In some embodiments, a covalent attachment is mediated through one or more linkers. In some embodiments, the linker is a cleavable linker . In certain embodiments, the cleavable linker comprises an amide, ester, or disulfide bond. For example, the linker may be an amino acid sequence that is cleavable by a cellular enzyme. In certain embodiments, the enzyme is a protease. In other embodiments, the enzyme is an esterase. In some embodiments, the enzyme is one that is more highly expressed in certain cell types than in other cell types. For example, the enzyme may be one that is more highly expressed in tumor cells than in non-tumor cells. Exemplary linkers and enzymes that cleave those linkers are presented below.

Cleavable Linkers

Linker Sequence	Enzyme(s) Targeting Linker
X^1-AGVF-X (SEQ ID NO: 256)	lysosomal thiol proteinases (see, *e.g.*, Duncan *et al.*, 1982, *Biosci. Rep.*, 2:1041-46; incorporated herein by reference)
X-GFLG-X	lysosomal cysteine proteinases (see, *e.g.*, Vasey *et al.*, *Clin. Canc.*

(SEQ ID NO: 257)	*Res.*, 1999, 5:83-94; incorporated herein by reference)
X-FK-X	Cathepsin B – ubiquitous, overexpressed in many solid tumors, such as breast cancer (see, *e.g.*, Dubowchik *et al.*, 2002, *Bioconjugate Chem.*, 13:855-69; incorporated herein by reference)
X-A*L-X	Cathepsin B – ubiquitous, overexpressed in many solid tumors, such as breast cancer (see, *e.g.*, Trouet *et al.*, 1982, *Proc. Natl. Acad. Sci., USA*, 79:626-29; incorporated herein by reference)
X-A*LA*L-X (SEQ ID NO: 258)	Cathepsin B – ubiquitous, overexpressed in many solid tumors (see, *e.g.*, Schmid *et al.*, 2007, *Bioconjugate Chem*, 18:702-16; incorporated herein by reference)
X-AL*AL*A-X (SEQ ID NO: 259)	Cathepsin D – ubiquitous (see, *e.g.*, Czerwinski *et al.*, 1998, *Proc. Natl. Acad. Sci., USA*, 95:11520-25; incorporated herein by reference)

[1] X denotes a supercharged protein or a functional effector protein to be delivered
* refers to observed cleavage site

[0075] To give but one particular example, a +36 GFP may be associated with a functional effector protein to be delivered by a cleavable linker, such as ALAL (SEQ ID NO: 254), to generate +36 GFP-$(GGS)_4$-ALAL-$(GGS)_4$-[functional effector protein X] (SEQ ID NO: 255).

[0076] In certain embodiments, the functional effector protein to be delivered is contacted with the supercharged protein to form a complex. In some embodiments, formation of complexes is carried out at or around pH 7. In some embodiments, formation of complexes is carried out at about pH 5, about pH 6, about pH 7, about pH 8, or about pH 9. Formation of complexes is typically carried out at a pH that does not negatively affect the function of the supercharged protein and/or the functional effector protein. In some embodiments, formation of complexes is carried out at room temperature. In some embodiments, formation of complexes is carried out at or around 37 ºC. In some embodiments, formation of complexes is carried out below 4 ºC, at about 4 ºC, at about 10 ºC, at about 15 ºC, at about 20 ºC, at about 25 ºC, at about 30 ºC, at about 35 ºC, at about 37 °C, at about 40 ºC, or higher than 40 ºC. Formation of complexes is typically carried out at a temperature that does not negatively affect the function of the supercharged protein and/or functional effector protein. In some embodiments, formation of complexes is carried out in serum-free medium. In some embodiments, formation of complexes is carried out in the presence of CO_2 (*e.g.*, about 1%, about 2%, about 3%, about 4%, about 5%, about 6%, or more).

[0077] In some embodiments, formation of complexes is carried out using concentrations of functional effector protein of about 100 nM. In some embodiments, formation of complexes is carried out using concentrations of functional effector protein of about 25 nM, about 50 nM, about 75 nM, about 90 nM, about 100 nM, about 110 nM, about 125 nM, about 150 nM, about 175 nM, or about 200 nM. In some embodiments, formation of complexes is carried out using concentrations of supercharged protein of about 40 nM. In some embodiments, formation of complexes is carried out using concentrations of supercharged protein of about 10 nM, about 20 nM, about 30 nM, about 40 nM, about 50 nM, about 60 nM, about 70 nM, about 80 nM, about 90 nM, or about 100 nM.

[0078] In some embodiments, formation of complexes is carried out under conditions of excess functional effector protein. In some embodiments, formation of complexes is carried out with ratios of functional effector protein:supercharged protein of about 20:1, about 10:1, about 9:1, about 8:1, about 7:1, about 6:1, about 5:1, about 4:1, about 3:1, about 2:1, or about 1:1. In some embodiments, formation of complexes is carried out with ratios of functional effector protein:supercharged protein of about 3:1. In some embodiments, formation of complexes is carried out with ratios of supercharged protein: functional effector protein of about 20:1, about 10:1, about 9:1, about 8:1, about 7:1, about 6:1, about 5:1, about 4:1, about 3:1, about 2:1, or about 1:1.

[0079] In some embodiments, formation of complexes is carried out by mixing supercharged protein with functional effector protein, and agitating the mixture (*e.g.*, by inversion). In some embodiments, formation of complexes is carried out by mixing supercharged protein with functional effector protein, and allowing the mixture to sit still. In some embodiments, the formation of the complex is carried out in the presence of a pharmaceutically acceptable carrier or excipient. In some embodiments, the complex is further combined with a pharmaceutically acceptable carrier or excipient. Exemplary excipients or carriers include water, solvents, lipids, proteins, peptides, endosomolytic agents (*e.g.*, chloroquine, pyrene butyric acid), small molecules, carbohydrates, buffers, natural polymers, synthetic polymers (*e.g.*, PLGA, polyurethane, polyesters, polycaprolactone, polyphosphazenes), pharmaceutical agents, *etc.*

[0080] In some embodiments, complexes comprising supercharged protein and functional effector protein may migrate more slowly in gel electrophoresis assays than either the supercharged protein alone or the functional effector protein alone.

Applications

[0081] The present invention provides compositions comprising supercharged proteins, naturally occurring or engineered, associated with functional effector proteins (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) to be delivered to a cell, as well as methods of using such compositions and uses of such compositions. In certain embodiments, compositions are provided comprising a Cas9 protein (*e.g.*, wherein the Cas9 protein is associated with a gRNA) and a cationic lipid. In certain embodiments, compositions are provided comprising a Cas9 protein (*e.g.*, wherein the Cas9 protein is associated with a gRNA) and a cationic polymer. The inventive compositions may be used to treat or prevent any disease that can benefit, *e.g.*, from the delivery of an agent to a cell. The inventive compositions may also be used to transfect or treat cells for research purposes.

[0082] In some embodiments, compositions in accordance with the invention may be used for research purposes, *e.g.*, to efficiently deliver functional effector proteins to cells in a research context. In some embodiments, compositions in accordance with the present invention may be used for therapeutic purposes. In some embodiments, compositions in accordance with the present invention may be used for treatment of any of a variety of diseases, disorders, and/or conditions, including, but not limited to, one or more of the following: autoimmune disorders (*e.g.*, diabetes, lupus, multiple sclerosis, psoriasis, rheumatoid arthritis); inflammatory disorders (*e.g.*, arthritis, pelvic inflammatory disease); infectious diseases (*e.g.*, viral infections (*e.g.*, HIV, HCV, RSV), bacterial infections, fungal infections, sepsis); neurological disorders (*e.g.* Alzheimer's disease, Huntington's disease; autism; Duchenne muscular dystrophy); cardiovascular disorders (*e.g.* atherosclerosis, hypercholesterolemia, thrombosis, clotting disorders, angiogenic disorders such as macular degeneration); proliferative disorders (*e.g.* cancer, benign neoplasms); respiratory disorders (*e.g.* chronic obstructive pulmonary disease); digestive disorders (*e.g.* inflammatory bowel disease, ulcers); musculoskeletal disorders (*e.g.* fibromyalgia, arthritis); endocrine, metabolic, and nutritional disorders (*e.g.* diabetes, osteoporosis); urological disorders (*e.g.* renal disease); psychological disorders (*e.g.* depression, schizophrenia); skin disorders (*e.g.* wounds, eczema); blood and lymphatic disorders (*e.g.* anemia, hemophilia); *etc.*

[0083] Compositions of the invention may be used in a clinical setting. For example, a supercharged protein may be associated with a functional effector protein that can be used for therapeutic applications. Such functional effector protein may be, for example, nucleases or

transcriptional activators. Other compositions comprising a Cas9 protein and a cationic lipid may also be used for therapeutic applications.

[0084] In some embodiments, the supercharged protein or functional effector protein associated with a supercharged protein includes a detectable label. These molecules can be used in detection, imaging, disease staging, diagnosis, or patient selection. Suitable labels include fluorescent, chemiluminescent, enzymatic labels, colorimetric, phosphorescent, density-based labels, *e.g.*, labels based on electron density, and in general contrast agents, and/or radioactive labels.

Pharmaceutical Compositions

[0085] The present invention provides compositions comprising supercharged proteins associated with at least one functional effector protein to be delivered, and in some embodiments are encapsulated by cationic lipids. Other compositions comprising a Cas9 protein and a cationic lipid are provided. Thus, the present invention provides pharmaceutical compositions comprising one or more supercharged proteins associated with a functional effector protein, and/or one or more functional effector proteins associated with a cationic lipid and/or cationic polymer, and one or more pharmaceutically acceptable excipients. Pharmaceutical compositions may optionally comprise one or more additional therapeutically active substances. In accordance with some embodiments, a method of administering pharmaceutical compositions comprising one or more supercharged proteins associated with a functional effector protein to be delivered to a subject in need thereof is provided. In some embodiments, compositions are administered to humans. For the purposes of the present disclosure, the phrase "active ingredient" generally refers to a Cas9 protein and/or supercharged protein associated with a functional effector protein, or to the functional effector protein to be delivered as described herein.

[0086] Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals of all sorts. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates;

mammals, including commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

[0087] Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit.

[0088] A pharmaceutical composition in accordance with the invention may be prepared, packaged, and/or sold in bulk, as a single unit dose, and/or as a plurality of single unit doses. As used herein, a "unit dose" is discrete amount of the pharmaceutical composition comprising a predetermined amount of the active ingredient. The amount of the active ingredient is generally equal to the dosage of the active ingredient which would be administered to a subject and/or a convenient fraction of such a dosage such as, for example, one-half or one-third of such a dosage.

[0089] Relative amounts of the active ingredient, the pharmaceutically acceptable excipient, and/or any additional ingredients in a pharmaceutical composition in accordance with the invention will vary, depending upon the identity, size, and/or condition of the subject treated and further depending upon the route by which the composition is to be administered. By way of example, the composition may comprise between 0.1% and 100% (w/w) active ingredient.

[0090] Pharmaceutical formulations may additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's *The Science and Practice of Pharmacy*, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated herein by reference) discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. Except insofar as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this invention.

[0091] In some embodiments, a pharmaceutically acceptable excipient is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% pure. In some embodiments, an excipient is approved for use in humans and for veterinary use. In some embodiments, an excipient is approved by United States Food and Drug Administration. In some embodiments, an excipient is pharmaceutical grade. In some embodiments, an excipient meets the standards of the United States Pharmacopoeia (USP), the European Pharmacopoeia (EP), the British Pharmacopoeia, and/or the International Pharmacopoeia.

[0092] Pharmaceutically acceptable excipients used in the manufacture of pharmaceutical compositions include, but are not limited to, inert diluents, dispersing and/or granulating agents, surface active agents and/or emulsifiers, disintegrating agents, binding agents, preservatives, buffering agents, lubricating agents, and/or oils. Such excipients may optionally be included in pharmaceutical formulations. Excipients such as cocoa butter and suppository waxes, coloring agents, coating agents, sweetening, flavoring, and/or perfuming agents can be present in the composition, according to the judgment of the formulator.

[0093] Exemplary diluents include, but are not limited to, calcium carbonate, sodium carbonate, calcium phosphate, dicalcium phosphate, calcium sulfate, calcium hydrogen phosphate, sodium phosphate lactose, sucrose, cellulose, microcrystalline cellulose, kaolin, mannitol, sorbitol, inositol, sodium chloride, dry starch, cornstarch, powdered sugar, *etc.*, and/or combinations thereof.

[0094] Exemplary granulating and/or dispersing agents include, but are not limited to, potato starch, corn starch, tapioca starch, sodium starch glycolate, clays, alginic acid, guar gum, citrus pulp, agar, bentonite, cellulose and wood products, natural sponge, cation-exchange resins, calcium carbonate, silicates, sodium carbonate, cross-linked poly(vinyl-pyrrolidone) (crospovidone), sodium carboxymethyl starch (sodium starch glycolate), carboxymethyl cellulose, cross-linked sodium carboxymethyl cellulose (croscarmellose), methylcellulose, pregelatinized starch (starch 1500), microcrystalline starch, water insoluble starch, calcium carboxymethyl cellulose, magnesium aluminum silicate (Veegum), sodium lauryl sulfate, quaternary ammonium compounds, *etc.*, and/or combinations thereof.

[0095] Exemplary surface active agents and/or emulsifiers include, but are not limited to, natural emulsifiers (*e.g.* acacia, agar, alginic acid, sodium alginate, tragacanth, chondrux, cholesterol, xanthan, pectin, gelatin, egg yolk, casein, wool fat, cholesterol, wax, and lecithin), colloidal clays (*e.g.* bentonite [aluminum silicate] and Veegum® [magnesium aluminum silicate]), long chain amino acid derivatives, high molecular weight alcohols (*e.g.* stearyl alcohol, cetyl alcohol, oleyl alcohol, triacetin monostearate, ethylene glycol distearate,

glyceryl monostearate, and propylene glycol monostearate, polyvinyl alcohol), carbomers (*e.g.* carboxy polymethylene, polyacrylic acid, acrylic acid polymer, and carboxyvinyl polymer), carrageenan, cellulosic derivatives (*e.g.* carboxymethylcellulose sodium, powdered cellulose, hydroxymethyl cellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, methylcellulose), sorbitan fatty acid esters (*e.g.* polyoxyethylene sorbitan monolaurate (Tween®20), polyoxyethylene sorbitan (Tween®60), polyoxyethylene sorbitan monooleate [Tween®80], sorbitan monopalmitate [Span®40], sorbitan monostearate [Span®60], sorbitan tristearate [Span®65], glyceryl monooleate, sorbitan monooleate [Span®80]), polyoxyethylene esters (*e.g.* polyoxyethylene monostearate [Myrj®45], polyoxyethylene hydrogenated castor oil, polyethoxylated castor oil, polyoxymethylene stearate, and Solutol®), sucrose fatty acid esters, polyethylene glycol fatty acid esters (*e.g.* Cremophor®), polyoxyethylene ethers, (*e.g.* polyoxyethylene lauryl ether (Brij®30)), poly(vinyl-pyrrolidone), diethylene glycol monolaurate, triethanolamine oleate, sodium oleate, potassium oleate, ethyl oleate, oleic acid, ethyl laurate, sodium lauryl sulfate, Pluronic®F 68, Poloxamer®188, cetrimonium bromide, cetylpyridinium chloride, benzalkonium chloride, docusate sodium, *etc.* and/or combinations thereof.

[0096] Exemplary binding agents include, but are not limited to, starch (*e.g.* cornstarch and starch paste); gelatin; sugars (*e.g.* sucrose, glucose, dextrose, dextrin, molasses, lactose, lactitol, mannitol,); natural and synthetic gums (*e.g.* acacia, sodium alginate, extract of Irish moss, panwar gum, ghatti gum, mucilage of isapol husks, carboxymethylcellulose, methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, microcrystalline cellulose, cellulose acetate, poly(vinyl-pyrrolidone), magnesium aluminum silicate (Veegum®), and larch arabogalactan); alginates; polyethylene oxide; polyethylene glycol; inorganic calcium salts; silicic acid; polymethacrylates; waxes; water; alcohol; *etc.*; and combinations thereof.

[0097] Exemplary preservatives may include, but are not limited to, antioxidants, chelating agents, antimicrobial preservatives, antifungal preservatives, alcohol preservatives, acidic preservatives, and/or other preservatives. Exemplary antioxidants include, but are not limited to, alpha tocopherol, ascorbic acid, acorbyl palmitate, butylated hydroxyanisole, butylated hydroxytoluene, monothioglycerol, potassium metabisulfite, propionic acid, propyl gallate, sodium ascorbate, sodium bisulfite, sodium metabisulfite, and/or sodium sulfite. Exemplary chelating agents include ethylenediaminetetraacetic acid (EDTA), citric acid monohydrate, disodium edetate, dipotassium edetate, edetic acid, fumaric acid, malic acid, phosphoric acid, sodium edetate, tartaric acid, and/or trisodium edetate. Exemplary

antimicrobial preservatives include, but are not limited to, benzalkonium chloride, benzethonium chloride, benzyl alcohol, bronopol, cetrimide, cetylpyridinium chloride, chlorhexidine, chlorobutanol, chlorocresol, chloroxylenol, cresol, ethyl alcohol, glycerin, hexetidine, imidurea, phenol, phenoxyethanol, phenylethyl alcohol, phenylmercuric nitrate, propylene glycol, and/or thimerosal. Exemplary antifungal preservatives include, but are not limited to, butyl paraben, methyl paraben, ethyl paraben, propyl paraben, benzoic acid, hydroxybenzoic acid, potassium benzoate, potassium sorbate, sodium benzoate, sodium propionate, and/or sorbic acid. Exemplary alcohol preservatives include, but are not limited to, ethanol, polyethylene glycol, phenol, phenolic compounds, bisphenol, chlorobutanol, hydroxybenzoate, and/or phenylethyl alcohol. Exemplary acidic preservatives include, but are not limited to, vitamin A, vitamin C, vitamin E, beta-carotene, citric acid, acetic acid, dehydroacetic acid, ascorbic acid, sorbic acid, and/or phytic acid. Other preservatives include, but are not limited to, tocopherol, tocopherol acetate, deteroxime mesylate, cetrimide, butylated hydroxyanisol (BHA), butylated hydroxytoluened (BHT), ethylenediamine, sodium lauryl sulfate (SLS), sodium lauryl ether sulfate (SLES), sodium bisulfite, sodium metabisulfite, potassium sulfite, potassium metabisulfite, Glydant Plus®, Phenonip®, methylparaben, Germall®115, Germaben®II, Neolone™, Kathon™, and/or Euxyl®.

[0098] Exemplary buffering agents include, but are not limited to, citrate buffer solutions, acetate buffer solutions, phosphate buffer solutions, ammonium chloride, calcium carbonate, calcium chloride, calcium citrate, calcium glubionate, calcium gluceptate, calcium gluconate, D-gluconic acid, calcium glycerophosphate, calcium lactate, propanoic acid, calcium levulinate, pentanoic acid, dibasic calcium phosphate, phosphoric acid, tribasic calcium phosphate, calcium hydroxide phosphate, potassium acetate, potassium chloride, potassium gluconate, potassium mixtures, dibasic potassium phosphate, monobasic potassium phosphate, potassium phosphate mixtures, sodium acetate, sodium bicarbonate, sodium chloride, sodium citrate, sodium lactate, dibasic sodium phosphate, monobasic sodium phosphate, sodium phosphate mixtures, tromethamine, magnesium hydroxide, aluminum hydroxide, alginic acid, pyrogen-free water, isotonic saline, Ringer's solution, ethyl alcohol, *etc.*, and/or combinations thereof.

[0099] Exemplary lubricating agents include, but are not limited to, magnesium stearate, calcium stearate, stearic acid, silica, talc, malt, glyceryl behanate, hydrogenated vegetable oils, polyethylene glycol, sodium benzoate, sodium acetate, sodium chloride, leucine, magnesium lauryl sulfate, sodium lauryl sulfate, *etc.*, and combinations thereof.

[00100] Exemplary oils include, but are not limited to, almond, apricot kernel, avocado, babassu, bergamot, black current seed, borage, cade, camomile, canola, caraway, carnauba, castor, cinnamon, cocoa butter, coconut, cod liver, coffee, corn, cotton seed, emu, eucalyptus, evening primrose, fish, flaxseed, geraniol, gourd, grape seed, hazel nut, hyssop, isopropyl myristate, jojoba, kukui nut, lavandin, lavender, lemon, litsea cubeba, macademia nut, mallow, mango seed, meadowfoam seed, mink, nutmeg, olive, orange, orange roughy, palm, palm kernel, peach kernel, peanut, poppy seed, pumpkin seed, rapeseed, rice bran, rosemary, safflower, sandalwood, sasquana, savoury, sea buckthorn, sesame, shea butter, silicone, soybean, sunflower, tea tree, thistle, tsubaki, vetiver, walnut, and wheat germ oils. Exemplary oils include, but are not limited to, butyl stearate, caprylic triglyceride, capric triglyceride, cyclomethicone, diethyl sebacate, dimethicone 360, isopropyl myristate, mineral oil, octyldodecanol, oleyl alcohol, silicone oil, and/or combinations thereof.

[00101] Liquid dosage forms for oral and parenteral administration include, but are not limited to, pharmaceutically acceptable emulsions, microemulsions, solutions, suspensions, syrups, and/or elixirs. In addition to active ingredients, liquid dosage forms may comprise inert diluents commonly used in the art such as, for example, water or other solvents, solubilizing agents and emulsifiers such as ethyl alcohol, isopropyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils (in particular, cottonseed, groundnut, corn, germ, olive, castor, and sesame oils), glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols and fatty acid esters of sorbitan, and mixtures thereof. Besides inert diluents, oral compositions can include adjuvants such as wetting agents, emulsifying and suspending agents, sweetening, flavoring, and/or perfuming agents. In certain embodiments for parenteral administration, compositions are mixed with solubilizing agents such as Cremophor®, alcohols, oils, modified oils, glycols, polysorbates, cyclodextrins, polymers, and/or combinations thereof.

[00102] Injectable preparations, for example, sterile injectable aqueous or oleaginous suspensions may be formulated according to the known art using suitable dispersing agents, wetting agents, and/or suspending agents. Sterile injectable preparations may be sterile injectable solutions, suspensions, and/or emulsions in nontoxic parenterally acceptable diluents and/or solvents, for example, as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, U.S.P., and isotonic sodium chloride solution. Sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including

synthetic mono- or diglycerides. Fatty acids such as oleic acid can be used in the preparation of injectables.

[00103] Injectable formulations can be sterilized, for example, by filtration through a bacterial-retaining filter, and/or by incorporating sterilizing agents in the form of sterile solid compositions which can be dissolved or dispersed in sterile water or other sterile injectable medium prior to use.

[00104] In order to prolong the effect of an active ingredient, it is often desirable to slow the absorption of the active ingredient from subcutaneous or intramuscular injection. This may be accomplished by the use of a liquid suspension of crystalline or amorphous material with poor water solubility. The rate of absorption of the drug then depends upon its rate of dissolution which, in turn, may depend upon crystal size and crystalline form. Alternatively, delayed absorption of a parenterally administered drug form is accomplished by dissolving or suspending the drug in an oil vehicle. Injectable depot forms are made by forming microencapsule matrices of the drug in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are prepared by entrapping the drug in liposomes or microemulsions which are compatible with body tissues.

[00105] Compositions for rectal or vaginal administration are typically suppositories which can be prepared by mixing compositions with suitable non-irritating excipients such as cocoa butter, polyethylene glycol or a suppository wax which are solid at ambient temperature but liquid at body temperature and therefore melt in the rectum or vaginal cavity and release the active ingredient.

[00106] Solid dosage forms for oral administration include capsules, tablets, pills, powders, and granules. In such solid dosage forms, an active ingredient is mixed with at least one inert, pharmaceutically acceptable excipient such as sodium citrate or dicalcium phosphate and/or fillers or extenders (*e.g.* starches, lactose, sucrose, glucose, mannitol, and silicic acid), binders (*e.g.* carboxymethylcellulose, alginates, gelatin, polyvinylpyrrolidinone, sucrose, and acacia), humectants (*e.g.* glycerol), disintegrating agents (*e.g.* agar, calcium carbonate, potato or tapioca starch, alginic acid, certain silicates, and sodium carbonate), solution retarding agents (*e.g.* paraffin), absorption accelerators (*e.g.* quaternary ammonium compounds), wetting agents (*e.g.* cetyl alcohol and glycerol monostearate), absorbents (*e.g.* kaolin and bentonite clay), and lubricants (*e.g.* talc, calcium stearate, magnesium stearate,

solid polyethylene glycols, sodium lauryl sulfate), and mixtures thereof. In the case of capsules, tablets and pills, the dosage form may comprise buffering agents.

[00107] Solid compositions of a similar type may be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polyethylene glycols and the like. Solid dosage forms of tablets, dragees, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings and other coatings well known in the pharmaceutical formulating art. They may optionally comprise opacifying agents and can be of a composition that they release the active ingredient(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed manner. Examples of embedding compositions which can be used include polymeric substances and waxes. Solid compositions of a similar type may be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polyethylene glycols and the like.

[00108] Dosage forms for topical and/or transdermal administration of a composition may include ointments, pastes, creams, lotions, gels, powders, solutions, sprays, inhalants and/or patches. Generally, an active ingredient is admixed under sterile conditions with a pharmaceutically acceptable excipient and/or any needed preservatives and/or buffers as may be required. Additionally, the present invention contemplates the use of transdermal patches, which often have the added advantage of providing controlled delivery of a compound to the body. Such dosage forms may be prepared, for example, by dissolving and/or dispensing the compound in the proper medium. Alternatively or additionally, rate may be controlled by either providing a rate controlling membrane and/or by dispersing the compound in a polymer matrix and/or gel.

[00109] Suitable devices for use in delivering intradermal pharmaceutical compositions described herein include short needle devices. Intradermal compositions may be administered by devices which limit the effective penetration length of a needle into the skin and functional equivalents thereof. Jet injection devices which deliver liquid compositions to the dermis via a liquid jet injector and/or via a needle which pierces the stratum corneum and produces a jet which reaches the dermis are suitable. Ballistic powder/particle delivery devices which use compressed gas to accelerate vaccine in powder form through the outer layers of the skin to the dermis are suitable. Alternatively or additionally, conventional syringes may be used in the classical mantoux method of intradermal administration.

[00110] Formulations suitable for topical administration include, but are not limited to, liquid and/or semi liquid preparations such as liniments, lotions, oil in water and/or water in

oil emulsions such as creams, ointments and/or pastes, and/or solutions and/or suspensions. Topically-administrable formulations may, for example, comprise from about 1% to about 10% (w/w) active ingredient, although the concentration of active ingredient may be as high as the solubility limit of the active ingredient in the solvent. Formulations for topical administration may further comprise one or more of the additional ingredients described herein.

[00111] A pharmaceutical composition may be prepared, packaged, and/or sold in a formulation suitable for pulmonary administration via the buccal cavity. Such a formulation may comprise dry particles which comprise the active ingredient and which have a diameter in the range from about 0.5 nm to about 7 nm or from about 1 nm to about 6 nm. Such compositions are conveniently in the form of dry powders for administration using a device comprising a dry powder reservoir to which a stream of propellant may be directed to disperse the powder and/or using a self-propelling solvent/powder dispensing container such as a device comprising the active ingredient dissolved and/or suspended in a low-boiling propellant in a sealed container. Such powders comprise particles wherein at least 98% of the particles by weight have a diameter greater than 0.5 nm and at least 95% of the particles by number have a diameter less than 7 nm. Alternatively, at least 95% of the particles by weight have a diameter greater than 1 nm and at least 90% of the particles by number have a diameter less than 6 nm. Dry powder compositions may include a solid fine powder diluent such as sugar and are conveniently provided in a unit dose form.

[00112] Low boiling propellants generally include liquid propellants having a boiling point of below 65 °F at atmospheric pressure. Generally the propellant may constitute 50% to 99.9% (w/w) of the composition, and active ingredient may constitute 0.1% to 20% (w/w) of the composition. A propellant may further comprise additional ingredients such as a liquid non-ionic and/or solid anionic surfactant and/or a solid diluent (which may have a particle size of the same order as particles comprising the active ingredient).

[00113] Pharmaceutical compositions formulated for pulmonary delivery may provide an active ingredient in the form of droplets of a solution and/or suspension. Such formulations may be prepared, packaged, and/or sold as aqueous and/or dilute alcoholic solutions and/or suspensions, optionally sterile, comprising active ingredient, and may conveniently be administered using any nebulization and/or atomization device. Such formulations may further comprise one or more additional ingredients including, but not limited to, a flavoring agent such as saccharin sodium, a volatile oil, a buffering agent, a surface active agent, and/or a preservative such as methylhydroxybenzoate. Droplets provided by this route of

administration may have an average diameter in the range from about 0.1 nm to about 200 nm.

[00114] Formulations described herein as being useful for pulmonary delivery are useful for intranasal delivery of a pharmaceutical composition. Another formulation suitable for intranasal administration is a coarse powder comprising the active ingredient and having an average particle from about 0.2 μm to 500 μm. Such a formulation is administered in the manner in which snuff is taken, *i.e.* by rapid inhalation through the nasal passage from a container of the powder held close to the nose.

[00115] Formulations suitable for nasal administration may, for example, comprise from about as little as 0.1% (w/w) and as much as 100% (w/w) of active ingredient, and may comprise one or more of the additional ingredients described herein. A pharmaceutical composition may be prepared, packaged, and/or sold in a formulation suitable for buccal administration. Such formulations may, for example, be in the form of tablets and/or lozenges made using conventional methods, and may, for example, 0.1% to 20% (w/w) active ingredient, the balance comprising an orally dissolvable and/or degradable composition and, optionally, one or more of the additional ingredients described herein. Alternately, formulations suitable for buccal administration may comprise a powder and/or an aerosolized and/or atomized solution and/or suspension comprising active ingredient. Such powdered, aerosolized, and/or aerosolized formulations, when dispersed, may have an average particle and/or droplet size in the range from about 0.1 nm to about 200 nm, and may further comprise one or more of any additional ingredients described herein.

[00116] A pharmaceutical composition may be prepared, packaged, and/or sold in a formulation suitable for ophthalmic administration. Such formulations may, for example, be in the form of eye drops including, for example, a 0.1/1.0% (w/w) solution and/or suspension of the active ingredient in an aqueous or oily liquid excipient. Such drops may further comprise buffering agents, salts, and/or one or more other of any additional ingredients described herein. Other opthalmically-administrable formulations which are useful include those which comprise the active ingredient in microcrystalline form and/or in a liposomal preparation. Ear drops and/or eye drops are contemplated as being within the scope of this invention.

[00117] General considerations in the formulation and/or manufacture of pharmaceutical agents may be found, for example, in *Remington: The Science and Practice of Pharmacy* 21st ed., Lippincott Williams & Wilkins, 2005 (incorporated herein by reference).

Administration

[00118] The present invention provides methods comprising administering compositions of supercharged proteins associated with functional effector proteins to a subject in need thereof. In some embodiments, methods of administering compositions comprising other functional effector proteins (*e.g.*, a Cas9 protein) and cationic lipid and/or cationic polymers are provided. Such compositions may be administered to a subject using any amount and any route of administration effective for preventing, treating, diagnosing, or imaging a disease, disorder, and/or condition. The exact amount required will vary from subject to subject, depending on the species, age, and general condition of the subject, the severity of the disease, the particular composition, its mode of administration, its mode of activity, and the like. Compositions in accordance with the invention are typically formulated in dosage unit form for ease of administration and uniformity of dosage. It will be understood, however, that the total daily usage of the compositions of the present invention will be decided by the attending physician within the scope of sound medical judgment. The specific therapeutically effective, prophylactially effective, or appropriate imaging dose level for any particular patient will depend upon a variety of factors including the disorder being treated and the severity of the disorder; the activity of the specific compound employed; the specific composition employed; the age, body weight, general health, sex and diet of the patient; the time of administration, route of administration, and rate of excretion of the specific compound employed; the duration of the treatment; drugs used in combination or coincidental with the specific compound employed; and like factors well known in the medical arts.

[00119] Compositions of supercharged proteins associated with functional effector proteins to be delivered as well as compositions comprising *e.g.*, a Cas9 protein and cationic lipid may be administered by any route. In some embodiments, such compositions are administered by one or more of a variety of routes, including oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, subcutaneous, intraventricular, transdermal, interdermal, rectal, intravaginal, intraperitoneal, topical (*e.g.*, by powders, ointments, creams, gels, lotions, and/or drops), mucosal, nasal, buccal, enteral, vitreal, intratumoral, sublingual; by intratracheal instillation, bronchial instillation, and/or inhalation; as an oral spray, nasal spray, and/or aerosol, and/or through a portal vein catheter. In some embodiments, supercharged proteins or complexes, and/or pharmaceutical, prophylactic, diagnostic, or imaging compositions thereof, are administered by systemic intravenous injection. In

specific embodiments, supercharged proteins or complexes and/or pharmaceutical, prophylactic, diagnostic, or imaging compositions thereof may be administered intravenously and/or orally. In specific embodiments, such compositions may be administered in a way which allows the functional effector protein to cross the blood-brain barrier, vascular barrier, or other epithelial barrier.

[00120] In certain embodiments, compositions in accordance with the invention may be administered at dosage levels sufficient to deliver an amount of functional effector protein of from about 0.0001 mg/kg to about 100 mg/kg, from about 0.01 mg/kg to about 50 mg/kg, from about 0.1 mg/kg to about 40 mg/kg, from about 0.5 mg/kg to about 30 mg/kg, from about 0.01 mg/kg to about 10 mg/kg, from about 0.1 mg/kg to about 10 mg/kg, or from about 1 mg/kg to about 25 mg/kg, of subject body weight per day, one or more times a day, to obtain the desired therapeutic, diagnostic, prophylactic, or imaging effect. The desired dosage may be delivered three times a day, two times a day, once a day, every other day, every third day, every week, every two weeks, every three weeks, or every four weeks. In certain embodiments, the desired dosage may be delivered using multiple administrations (*e.g.*, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, or more administrations).

[00121] Compositions comprising supercharged proteins associated with functional effector proteins may be administered in combination with one or more other therapeutic, prophylactic, diagnostic, or imaging agents. By "in combination with," it is not intended to imply that the agents must be administered at the same time and/or formulated for delivery together, although these methods of delivery are within the scope of the invention. Compositions can be administered concurrently with, prior to, or subsequent to, one or more other desired therapeutics or medical procedures. In general, each agent will be administered at a dose and/or on a time schedule determined for that agent. In some embodiments, the invention encompasses the delivery of pharmaceutical, prophylactic, diagnostic, or imaging compositions in combination with agents that may improve their bioavailability, reduce and/or modify their metabolism, inhibit their excretion, and/or modify their distribution within the body.

Kits

[00122] The invention provides a variety of kits for conveniently and/or effectively carrying out methods of the present invention. Typically kits will comprise sufficient amounts and/or numbers of components to allow a user to perform multiple treatments of a

subject(s) and/or to perform multiple experiments. In some embodiments, kits comprise one or more of (i) a supercharged protein, as described herein; (ii) a functional effector protein to be delivered; (ii) a cationic lipid and/or cationic polymer; and (iv) instructions for formulating a composition comprising the functional protein associated to the supercharged protein. In some embodiments, the kits comprise a Cas9 protein and a cationic lipid. In some embodiments, kits comprise a nucleic acid encoding for the supercharged protein and/or the functional protein to be delivered. In some embodiments, the kit comprises a cloning vector encoding a supercharged protein and a cloning site allowing the in-frame cloning of a functional effector protein to generate a fusion protein. In some embodiments, kits comprise a pharmaceutical composition provided herein comprising a supercharged protein associated with a functional effector protein; a syringe, needle, or applicator for administration of the pharmaceutical composition to a subject; and instructions for administration of the pharmaceutical composition to the subject.

[00249] These and other aspects of the present invention will be further appreciated upon consideration of the following Examples, which are intended to illustrate certain particular embodiments of the invention but are not intended to limit its scope, as defined by the claims.

EXAMPLES

Example 1: Delivery of TALE Activators fused to supercharged GFP

[00250] A major target for reprogramming fibroblast cell fate towards brown or white adipocyte cell fate lies in the switch from White Adipose Tissue (WAT) to Brown Adipose Tissue (BAT), which is governed by expression of PRDM16 and PPARγ. Robust TALE transcriptional activators fused to a +36 GFP were engineered that target PPARγ and PRDM16 genomic sequences in fibroblasts. Fusion proteins were purified using a heparin column and/or an SEC and gels show a single band at 130 kD The modulation of expression and effect on cellular phenotype after delivery of the TALE activators was compared to the modulation after viral delivery of a PPARγ cDNA followed by 7-day treatment with adipogenesis cocktail. It was observed that adipocytes formed upon treatment with +36 GFP TALEPRDM16 fusion. Expression of white adipose tissue marker genes was detected after delivery of supercharged PRDM16 TALE activators.

[00251] A one-time supercharged protein-mediated delivery of a TALE activator for PPARγ was found to induce expression of white-fat genes and to differentiate fibroblasts into

white-fat cells. Supercharged protein-mediated delivery of both a PPARγ and PRDM16 TALE activator induced the differentiation of fat cells with increased expression of brown-fat markers such as PRDM16, cox8b, elovl3, and cidea as well as a small increase in thermogenic gene expression markers PGC1a and UCP1.

[00252] An Aurein peptide was fused to the N-terminus of the +36GFP-TALE-activator fusion protein. Delivery of Aurein +36 GFP TALE purified by heparin column was observed by detecting fluorescence in nucleus of the treated cells.

[00253] Figure 1 shows a schematic of macromolecular delivery into mammalian cells. Figure 2 shows an overview of the switch from White Adipose Tissue (WAT) to Brown Adipose Tissue (BAT). Figure 3 shows a schematic of supercharged delivery platforms to deliver TALE activators programmed to target PPARγ or PRDM16. Figure 4 shows a schematic of a fusion protein comprising a +36 GFP fusion, an 18.5 mer TALE domain, and a VP64 activation domain. Figure 5 shows expression and purification of the +36 GFP-TALE activator-fusion protein. Figure 6 shows testing assays for activation of fat cell regulator genes upon delivery of +36 GFP PPARγ and PRDM16 TALE activator fusion proteins.

[00254] Figure 7 shows delivery efficacy of +36 GFP TALE activator fusion proteins at different concentrations. Figure 8 shows a comparison of delivery efficacy of two different +36 GFP-PRDM16 TALE fusion proteins in NIH 3T3 cells. Figure 9 shows PPARγ gene expression after delivery of PPARγ-TALE activator fusion and comparison to various controls. Figure 10 shows PRDM16 gene expression after delivery of RDM16-TALE activator fusion and comparison to various controls. Figure 11 shows moderate TALE activity observed in the presence of serum.

[00255] Figure 12 shows a validation of viral delivery of PPARγ followed by 7-day treatment with adipogenesis cocktail. Figure 13 shows a schematic of an assay for programming fibroblasts into WAT and BAT. Figure 14 shows adipocyte formation observed upon treatment with +36 GFP TALE activator fusion protein. Figure 15 shows staining of various treatments after 7 days with LipidTOX red, demonstrating formation of adipocytes after viral delivery as well as after delivery of supercharged PPARγ TALE activator fusion protein. Figure 16 shows staining of cells after various treatments after 7 days with LipidTOX red, demonstrating formation of adipocytes after viral delivery as well as after delivery of supercharged PPARγ TALE activator fusion protein. Figure 17 shows expression of WAT biomarker genes after viral delivery as well as after delivery of supercharged PPARγ TALE activator fusion protein.

Example 2: In vivo delivery of TALE Activators fused to supercharged GFP

[00256] NIH 3T3 cells were grown to 70-90% confluence and treated with 1 μM or between 0.5-5 μM of +36 GFP PPARγ TALE and/or +36 GFP PRDM16 TALE fusion protein in DMEM without serum. A serum-free medium was chosen, because serum can decrease the effectiveness of protein-based delivery. Cells were incubated with the respective fusion protein solution for 4 hours before the media was removed and full DMEM containing serum was added back to cells. Control cells were infected with a viral contruct encoding PPARγ or PRDM16 in order to serve as a positive control for expression of WAT and BAT genes according to known protocols (see, *e.g.*, Seale *et al.* 2008 *Nature* 454, 961-967, the entire contents of which are incorporated herein by reference). Once all cells reached 100% confluence an adipogenesis cocktail containing isobutylmethylxanthine, insulin, rosiglitazone, dexamethosone, T3, and indomethacin was added to the cells and replaced 48 hours later with a form of the cocktail containing only insulin, T3, and rosiglitazone. At 48 hours after this second replacement of cocktail another dosage of T3, insulin, and rosiglitazone was added to the cells. The next day, which is now one week from the start of the experiment, cells were harvested with TRIzol, total RNA was extracted, and qRT-PCR was performed to measure gene expression levels of PPARγ, PRDM16, and other brown fat marker genes such as UCP1, PGC1a, Elovl3, and Cidea.

[00257] Figure 18 shows delivery of supercharged PRDM16 TALE activator fusion proteins to induce brown-fat adipocytes *in vivo*. Robust adipocyte formation was observed after viral delivery of PPARγ and PRDM16 and also after delivery of supercharged TALE activator protein fusions. Figure 19 shows a comparison of supercharged (TALE) and viral delivery of PPARγ and PRDM16 to cells. The figure shows TALE/TALE, viral/TALE, and viral/viral-induced expression of brown fat markers by expression of PPARγ and PRDM16. Figure 20 shows RT-qPCR assessments consistent with fat cell differentiation, which were also observed by LipidTOX staining.

Example 3: Delivery of TALE Activators complexed with supercharged GFP

[00258] In order to improve delivery efficacy, protein complexes in which the functional protein was non-covalently associated with the supercharged protein were generated and administered to cells. Figure 21. shows that delivery of functional TALE activator fusion proteins as complexes with +36 GFP improves TALE activator activity after delivery. Figure 22 shows PRDM16 gene expression after TALE activator fusion delivery

either as a fusion (+36GFP PRDM16 TALE-3) or a complex (+36GFP + PRDM16 TALE-3) with +36GFP. It was observed that delivery of complexes tended to increase TALE activator activity.

Example 4: Effect of Aurein fusions on delivery efficacy

[00259] Figure 23 shows the effect of an N-terminal Aurein peptide fusion to +36GFP on PRDM16 gene expression after TALE activator fusion delivery (either as a fusion or a complex with +36GFP). The Aurein peptide was fused to the N-terminus of the GFP-TALE construct via a GGS(9) (SEQ ID NO: 252) linker, resulting in an Aurein peptide-GGS(9) linker-(+)36 GFP protein-GGS(9) linker -PRDM16 TALE-3 fusion protein. The protein was purified using size exclusion chromatography.

Example 5: Delivery of TALE Activators complexed with supercharged GFP or cationic lipids

[00260] Figure 24 shows PRDM16 gene expression after TALE PRDM16 activator protein delivery either as a fusion with +36 GFP (+36GFP PRDM16 TALE-3), a complex with +36 GFP (+36GFP + PRDM16 TALE-3), or a complex with Lipofectamine LTX, for which an increase in gene expression was observed.

Example 6: Delivery of Cas9 fused to supercharged GFP

[00261] Supercharged delivery of Cas9 into mammalian cells would allow the application of powerful RNA-programmable nuclease technology in cells without the drawbacks of prior delivery methods. To this end, a Cas9 fusion with +36GFP was generated, using an ALAL linker. Figure 25 shows a schematic of the supercharged fusion protein with Cas9. Figure 26 shows the purification of wild-type Cas9 protein and Cas9 fusion proteins with +36GFP and Aurein-GGS9. The fusion protein is administered to cells in the same manner as the TALE activator fusion proteins above. The Cas9, once delivered to the cells, binds and cleaves its target site in the cellular genome. Nuclease activity in the target cells is detected via a suitable assay, *e.g.*, via southern blot or sequencing assay.

***Example 7: Efficient Delivery of Genome Editing Proteins* In Vitro *and* In Vivo**

[00262] Efficient intracellular delivery of proteins to the nucleus or cytoplasm is needed to fully realize the potential of protein therapeutics including genome-editing agents. Current methods of protein delivery often suffer from low tolerance for serum proteins, poor

endosomal escape, and limited *in vivo* efficacy. As demonstrated in this Example, common cationic lipid reagents originally developed for nucleic acid transfection can potently deliver proteins that are fused to negatively supercharged proteins, that contain natural anionic domains, or that natively bind to anionic nucleic acids. This approach mediates the functional delivery of Cre recombinase, TALE- and Cas9-based transcriptional activators, and Cas9:sgRNA nuclease complexes into cultured human cells at low nanomolar concentrations in media containing 10% serum. Lipid-based delivery can be > 1,000-fold more potent than cationic protein delivery strategies. Delivery of Cas9:sgRNA complexes resulted in genome modification efficiencies as high as 80% with substantially higher specificity compared to standard DNA transfection, likely due to the transient nature of delivered Cas9:sgRNA complexes. This approach also mediated efficient delivery of Cre recombinase and Cas9:sgRNA complexes into the mouse inner ear in vivo, achieving up to 90% Cre-mediated recombination and 20% Cas9-mediated genome modification in the targeted hair-cell population.

Materials and Methods

Construction of Cas9, Cre, and TALE fusion and sgRNA expression plasmids.

[00263] Sequences of all constructs used in this paper are listed below or provided elsewhere in the specification. All protein constructs were generated from previously reported plasmids for protein of interest cloned into a pET29a expression plasmid.

Expression and purification of *S. pyogenes* Cas9 and other proteins.

[00264] *E. coli* BL21 STAR (DE3) competent cells (Life Technologies) were transformed with pMJ806[47] encoding the *S. pyogenes* Cas9 fused to an N-terminal 10xHis-tag/maltose binding protein. The resulting expression strain was inoculated in Luria-Bertani (LB) broth containing 100 μg/mL of ampicillin at 37 °C overnight. The cells were diluted 1:100 into the same growth medium and grown at 37 °C to OD_{600} = ~0.6. The culture was incubated at 20 °C for 30 min, and isopropyl β-D-1- thiogalactopyranoside (IPTG) was added at 0.5 mM to induce Cas9 expression. After approximately 16 hours, the cells were collected by centrifugation at 8,000 *g* and resuspended in lysis buffer (50 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl, pH 8.0, 1 M NaCl, 20 % glycerol, 10 mM tris(2-carboxyethyl)phosphine (TCEP)). The cells were lysed by sonication (1 sec pulse-on, 1 sec pulse-off for 15 minutes total at 6 W output) and the soluble lysate was obtained by centrifugation at 20,000 *g* for 30 minutes.

[00265] The cell lysate was incubated with His-Pur nickel-nitriloacetic acid (nickel-NTA) resin (Thermo Scientific) at 4 °C for 30 minutes to capture His-tagged Cas9. The resin was transferred to a 20-mL column and washed with 20 column volumes of lysis buffer. Cas9 was eluted in 50 mM Tris-HCl (pH 8), 0.1 M NaCl, 20 % glycerol, 10 mM TCEP, and 300 mM imidazole, and concentrated by Amicon ultra centrifugal filter (Millipore, 100-kDa molecular weight cut-off) to ~50 mg/mL. The 6xHis tag and maltose-binding protein were removed by TEV protease treatment at 4 °C for 20 hours and captured by a second Ni-affinity purification step. The eluent, containing Cas9, was injected into a HiTrap SP HP column (GE Healthcare) in purification buffer containing 50 mM Tris-HCl (pH 8), 0.1 M NaCl, 20 % glycerol, and 10 mM TCEP. Cas9 was eluted with purification buffer containing a linear NaCl gradient from 0.1 M to 1 M over five column volumes. The eluted fractions containing Cas9 were concentrated down to a concentration of 200 μM as quantified by Bicinchoninic acid assay (BCA) (Pierce Biotechnology), snap-frozen in liquid nitrogen, and stored in aliquots at -80 °C. All other proteins were purified by this method but without TEV cleavage step and proteins containing (-30)GFP were purified by anion exchange using a Hi-Trap Q HP anion exchange column (GE Healthcare) using the same purification protocol.

***In vitro* transcription of sgRNAs.**

[00266] Linear DNA fragments containing the T7 promoter binding site followed by the 20-bp sgRNA target sequence were transcribed *in vitro* using the T7 High Yield RNA Synthesis Kit (NEB) according to the manufacturer's instructions. *In vitro* transcribed RNA was precipitated with ethanol and purified by gel electrophoresis on a Criterion 10% polyacrylamide TBE-Urea gel (Bio-Rad). Excised gel fragments were extracted in 420 μL of 300 mM NaCl overnight on a rocking surface at 4 °C. Gel-purified sgRNA was precipitated with ethanol and redissolved in water and sgRNA concentration was finally quantified by UV absorbance and snap-frozen at -80 °C.

Plasmid transfection.

[00267] Plasmid DNA was transfected using Lipofectamine 2000 (Life Technologies) according the manufacturer's protocol. For TALE activator plasmids, 300 ng of DNA was transfected, and for the activator synergy experiments 60 ng of each of five plasmids was pooled and transfected. For Cas9 nuclease delivery experiments, linear DNA PCR products expressing sgRNAs were used in transfection experiments targeting genomic sites in CLTA,

EMX, VEGF, and GFP (sgRNA GFP g1, GFP g3, GFP g5, and GFP g7 for nickase studies). Linear DNA PCR products were generated using plasmid containing the U6 promoter as template and forward primers bearing the U6 promoter upstream sequence and reverse primers containing U6 downstream sequence followed by the sgRNA sequence (20-bp sequence unique to each target plus constant sgRNA backbone architecture sequence). sgRNAs expressed from linear DNA templates contained at least two 5' guanosines to match *in vitro* transcribed sgRNAs that required these bases for T7 transcription. Primer sequences and PCR conditions are listed below. For dCas9 activator experiments, 700 ng of Cas9 or dCas9-VP64 plasmid DNA was co-transfected with 250 ng of the appropriate sgRNA expression plasmid. For activator synergy experiments 50 ng of DNA from each of the six sgRNA was pooled and co-transfected with 700 ng of dCas9-VP64 plasmid.

Delivery of transcription factor proteins complexed with cationic lipids in cell culture.

[00268] A more in-depth description of the delivery of genome-editing proteins both *in vitro* and *in vivo* can be found below. Briefly, cultured cells were plated in 48-well format (250 µL volume) in Dulbecco's Modified Eagle's Media plus GlutaMAX (Life Technologies, Carlsbad, CA) with 10% FBS ("full serum media") and antibiotics at a cell density necessary to reach ~70% confluence the next day. Full serum media was replaced with the same media but containing no antibiotics one hour before delivery. Delivery of Cre and TALE proteins was performed by combining 1 nM to 1 µM protein (in 275 µL final volume) with 0.5-1.5 µL of commercially available cationic lipids in 25 µL OPTIMEM media (Life Technologies, Carlsbad, CA) according to the manufacturer's protocol for normal plasmid transfection, including incubation time. For Cas9 delivery *in vitro*, transcribed sgRNA was incubated with Cas9 protein for 5 min before complexing with the cationic lipid reagent. 25 µL lipid complexes in OPTIMEM media were added to cells and media was replaced 12-16 hours later fresh media unless otherwise noted. Cells were assayed for recombination 48 hours after delivery, for gene activation either 4 or 16 hours after delivery, and for gene modification 72 hours after delivery.

T7 endonuclease I assay to detect genomic modifications.

[00269] U2OS-EGFP cells or HEK293T cells were transfected with Cas9 expression and sgRNA expression plasmids or linear DNA PCR products as described above or treated with only Cas9 protein, only *in vitro* transcribed sgRNA, or only RNAiMAX. Genomic DNA was isolated from cells 2 days after transfection using the DNAdvance Kit (Agencourt)

following the manufacturer's instructions. 200 ng of genomic DNA was used as template in PCR reactions to amplify the targeted genomic loci with flanking survey primer pairs specified below. PCR products were purified with a QIAquick PCR Purification Kit (Qiagen) and quantified with Quant-iT™ PicoGreen ® dsDNA Kit (Life Technologies). 250ng of purified PCR DNA was combined with 2 μL of NEBuffer 2 (NEB) in a total volume of 19 μL and denatured then re-annealed with thermocycling at 95 °C for 5 minutes, 95 to 85 °C at 2 °C/s; 85 to 20 °C at 0.2 °C/s. The re-annealed DNA was incubated with 1 μl of T7 Endonuclease I (10 U/μl, NEB) at 37 °C for 15 minutes. 10 μL of 50 % glycerol was added to the T7 Endonuclease reaction and 12 μL was analyzed on a 5 % TBE 18-well Criterion PAGE gel (Bio-Rad) electrophoresed for 30 minutes at 200 V, then stained with 1x SYBR Gold (Life Technologies) for 30 min. Cas9-induced cleavage bands and the uncleaved band were visualized on an AlphaImager HP (Alpha Innotech) and quantified using ImageJ software[54]. The peak intensities of the cleaved bands were divided by the total intensity of all bands (uncleaved + cleaved bands) to determine the fraction cleaved which was used to estimate gene modification levels as previously described.[46] For each sample, transfections and subsequent modification measurements were performed in triplicate on different days.

Stem cell culture and delivery.

[00270] Mouse embryonic stem cell (ES) line Tau-GFP containing a permanent GFP gene insertion was cultured in DMEM with 15% FBS (Gibco), 100 mM MEM nonessential amino acids (Gibco), 0.55 mM 2-mercaptoethanol, and leukemia inhibitory factor (1,000 units/ml; Chemicon). After 5 days floating spheres were formed that exhibited GFP fluorescence. Complexes of Cas9:sgRNA and RNAiMAX were added to the culture containing the floating spheres for 16 hours. After Cas9:sgRNA treatment, the cells were cultured in the above media for 3 days. The floating spheres were treated with trypsin for 5 minutes then passed through a 70 μm filter to collect single cells. The cells were cultured on laminin-coated slides in DMEM/F12 (1:1) supplemented with 1xN2, 1xB27, penicillin-streptomycin (100 μg/mL) and 10% FBS for two days before labeling. Immunohistochemistry was performed using an anti-GFP antibody (#ab13970, Abcam) to assess GFP expression. To quantify the number of GFP-negative cells, we counted the total number of GFP-positive and GFP-negative cells from three representative visual fields at 20X magnification, and calculated the average efficiency. Three independent experiments were performed for each condition.

Microinjection of proteins to mouse inner ear.

[00271] P0 floxP-tdTomato mice were used for (–30)GFP-Cre injection and P2 Atoh1-GFP mice were used for Cas9:sgRNA injection. Animals were used under protocols approved by the Massachusetts Eye & Ear Infirmary ALCUC committee. Mice were anesthetized by lowering their temperature on ice. Cochleostomies were performed by making an incision behind the ear to expose the otic bulla. Glass micropipettes held by a micromanipulator were used to deliver the complex into the scala media, which allows access to inner ear hair cells. For delivery of (-30)GFP-Cre, 3 μL of 45 μM protein was mixed with 3 μL of either RNAiMAX or Lipofectamine 2000 and incubated at room temperature for 30 minutes prior to injection. Four mice were injected per treatment group. For delivery of Cas9:sgRNA complexes, 1 μL of 200 μM Cas9 protein was mixed with 2 μL of 100 μM sgRNA and incubated for 5 minutes at room temperature before mixing with 3 μL of either RNAiMAX or Lipofectamine 2000 and incubating for an additional 30 minutes prior to injection. Three mice were injected per treatment group. The total delivery volume for every injection was 0.3 μL per cochlea and the release was controlled by a micromanipulator at the speed of 32 nL/sec.

Immunohistochemistry and quantification.

[00272] 5-10 days after injection, the mice were sacrificed and cochlea were harvested by standard protocols.[55] For immunohistochemistry, antibodies against hair-cell markers (Myo7a and Esp) and supporting cells (Sox2) were used following a previously described protocol.[55] To quantify the number of tdTomato positive cells after (–30)GFP-Cre or GFP negative cells after Cas9:sgRNA delivery, we counted the total number of outer hair cells in a region spanning 200 μm around the site of injection in the base turn of the cochlea. The efficiency of (–30)GFP-Cre-induced recombination or Cas9:sgRNA-induced genome modification was calculated as the percentage of outer hair cells that expressed tdTomato or that lost GFP expression.

High-throughput DNA sequencing of genome modifications.

[00273] HEK293T cells were either transfected with Cas9 and sgRNA expression plasmids or linear DNA PCR products or treated with 50 nM Cas9 protein, 250 nM purified sgRNA, and cationic lipids as described earlier for Cas9 protein delivery to U2OS-EGFP

reporter cells. For plasmid-based transfection experiments, 700 ng of Cas9 expression plasmid plus 250 ng of sgRNA plasmid or 50 ng of a linear DNA PCR product expressing sgRNA for targeting either the EMX1, CLTA2, or VEGF locus were transfected with Lipofectamine 2000 (Life Technologies) and cells were isolated 2 days later. For protein delivery experiments *in vivo*, ~30 mg of mouse tissue was isolated as previously described[55] from anesthetized mice and genomic DNA was extracted using the Agencourt DNAAdvance Genomic DNA Isolation Kit (Beckman Coulter). For cell culture experiments genomic DNA was isolated as described above. 150 ng of genomic DNA was used as template to amplify by PCR the on-target and off-target genomic sites with flanking HTS primer pairs specified below. Relative amounts of crude PCR products were quantified by gel electrophoresis and samples treated with different sgRNA pairs or Cas9 nuclease types were separately pooled in equimolar concentrations before purification with the QIAquick PCR Purification Kit (Qiagen). Approximately150 ng of pooled DNA was electrophoresed using a 5% TBE 18-well Criterion PAGE gel (BioRad) for 30 min at 200 V and DNAs ~125 bp to ~300 bp in length were isolated and purified by QIAquick PCR Purification Kit (Qiagen). Purified DNA was amplified by PCR with primers containing sequencing adapters, purified, and sequenced on a MiSeq high-throughput DNA sequencer (Illumina) as previously described.[47]

Data Analysis

[00274] Illumina sequencing reads were filtered and parsed with scripts written in Unix Bash as outlined below. Sample sizes for sequencing experiments were maximized (within practical experimental considerations) to ensure greatest power to detect effects. Statistical analyses for Cas9-modified genomic sites (Table 2) were performed as previously described[56] with multiple comparison correction using the Bonferroni method.

[00275] The following is a list of upstream and downstream flanking sequences for each genomic target site.

Target Site	Downstream genomic sequence	Upstream genomic sequence
EMX_On	GGCCTGCTTCGTGGCAATGC (SEQ ID NO:119)	ACCTGGGCCAGGGAGGGAGG (SEQ ID NO:120)
EMX_Off1	CTCACTTAGACTTTCTCTCC (SEQ ID NO:121)	CTCGGAGTCTAGCTCCTGCA (SEQ ID NO:122)
EMX_Off2	TGGCCCCAGTCTCTCTTCTA (SEQ ID NO:123)	CAGCCTCTGAACAGCTCCCG (SEQ ID NO:124)

EMX_Off3	TGACTTGGCCTTTGTAGGAA (SEQ ID NO:125)	GAGGCTACTGAAACATAAGT (SEQ ID NO:126)
EMX_Off4	TGCTACCTGTACATCTGCAC (SEQ ID NO:127)	CATCAATGATTGGGCATTTC (SEQ ID NO:128)
VEG_On	ACTCCAGTCCCAAATATGTA (SEQ ID NO:129)	ACTAGGGGGCGCTCGGCCAC (SEQ ID NO:130)
VEG_Off1	CTGAGTCAACTGTAAGCATT (SEQ ID NO:131)	GGCCAGGTGCAGTGATTCAT (SEQ ID NO:132)
VEG_Off2	TCGTGTCATCTTGTTTGTGC (SEQ ID NO:133)	GGCAGAGCCCAGCGGACACT (SEQ ID NO:134)
VEG_Off3	CAAGGTGAGCCTGGGTCTGT (SEQ ID NO:135)	ATCACTGCCCAAGAAGTGCA (SEQ ID NO:136)
VEG_Off4	TTGTAGGATGTTTAGCAGCA (SEQ ID NO:137)	ACTTGCTCTCTTTAGAGAAC (SEQ ID NO:138)
CLT2_On	CTCAAGCAGGCCCCGCTGGT (SEQ ID NO:139)	TTTTGGACCAAACCTTTTTG (SEQ ID NO:140)
CLT2_Off1	TGAGGTTATTTGTCCATTGT (SEQ ID NO:141)	TAAGGGGAGTATTTACACCA (SEQ ID NO:142)
CLT2_Off2	TCAAGAGCAGAAAATGTGAC (SEQ ID NO:143)	CTTGCAGGGACCTTCTGATT (SEQ ID NO:144)
CLT2_Off3	TGTGTGTAGGACTAAACTCT (SEQ ID NO:145)	GATAGCAGTATGACCTTGGG (SEQ ID NO:146)
EGFP	AGCGTGTCCGGCGAGGGCGA (SEQ ID NO:147)	AGCGTGTCCGGCGAGGGCGA (SEQ ID NO:148)
MusEMX	CAGAATCGGAGGACAAAATACAAAC (SEQ ID NO:149)	ACGAAGCAGGCCAACGGGGAGGACA (SEQ ID NO:150)

Primers used for generating PCR products to serve as substrates for T7 transcription of sgRNAs.

[00276] T7_gRNA-Rev was used in all cases. DNA template used was EGFP sgRNA plasmid as noted above. NTF3 and VEGF sgRNAs for dCas9-VP64 activator experiments were reported previously (Maeder *et al.*, CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods*. 2013; 10, 977–979).

T7_EGFP1-Fwd TAA TAC GAC TCA CTA TA GGGCACGGGCAGCTTGCCGG (SEQ ID NO:151)

T7-GFP g1-Fwd TAA TAC GAC TCA CTA TA GGCCTCGAACTTCACCTCGGCG GAAAGGACGAAACACC (SEQ ID NO:152)

T7-GFP g5-Fwd TAA TAC GAC TCA CTA TA GGCTGAAGGGCATCGACTTCA

GAAAGGACGAAACACC (SEQ ID NO:153)

T7-GFP g3-Fwd TAA TAC GAC TCA CTA TA GGCAGCTCGATGCGGTTCACCA GAAAGGACGAAACACC (SEQ ID NO:154)

T7-GFP g7-Fwd TAA TAC GAC TCA CTA TA GGCAAGGAGGACGGCAACATCC GAAAGGACGAAACACC (SEQ ID NO:155)

T7-EMX-Fwd TAA TAC GAC TCA CTA TA GGAGTCCGAGCAGAAGAAGAA GAAAGGACGAAACACC (SEQ ID NO:156)

T7-VEG-Fwd TAA TAC GAC TCA CTA TA GGGGTGGGGGGAGTTTGCTCC GAAAGGACGAAACACC (SEQ ID NO:157)

T7-CLT2-Fwd TAA TAC GAC TCA CTA TA GGCAGATGTAGTGTTTCCACA GAAAGGACGAAACACC (SEQ ID NO:158)

T7_gRNA-Rev AAAAAAAGCACCGACTCGGTG (SEQ ID NO:159)

Primers for generating linear DNA PCR product for transfection.

[00277] PCR extension at (72 °C, 3 min) on plasmid containing U6 promoter as template with PCR_sgRNA-fwd1, PCR_sgRNA-rev2 and appropriate PCR_sgRNA primers listed below.

PCR_gRNA-fwd1 CTGTACAAAAAAGCAGGCTTTA (SEQ ID NO:160)

PCR_gRNA-rev2 AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGG ACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC (SEQ ID NO:161)

PCR-G-GFP1 GAAAGGACGAAACACC GGCCTCGAACTTCACCTCGGCGGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:162)

PCR-G-GFP3 GAAAGGACGAAACACC GGCAGCTCGATGCGGTTCACCAGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:163)

PCR-G-GFP5 GAAAGGACGAAACACC GGCTGAAGGGCATCGACTTCAGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:164)

PCR-G-GFP7 GAAAGGACGAAACACC GGCAAGGAGGACGGCAACATCCGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:165)

PCR-G-CLT2 GAAAGGACGAAACACC GGCAGATGTAGTGTTTCCACAGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:166)

PCR-G-EMX GAAAGGACGAAACACC GGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:167)

PCR-G-VEG	GAAAGGACGAAACACC GGGGTGGGGGGAGTTTGCTCCGTTTTAGAGCTAGAAATAGCAA (SEQ ID NO:168)

Primers for performing T7 endonuclease I DNA cleavage assay.

Survey_GFP-fwd	TACGGCAAGCTGACCCTGAA (SEQ ID NO:169)
Survey_GFP-rev	GTCCATGCCGAGAGTGATCC (SEQ ID NO:170)
Survey_CLTA-fwd	GCCAGGGGCTGTTATCTTGG (SEQ ID NO:171)
Survey_CLTA-rev	ATGCACAGAAGCACAGGTTGA (SEQ ID NO:172)
Survey_EMX-fwd	CTGTGTCCTCTTCCTGCCCT (SEQ ID NO:173)
Survey_EMX-rev	CTCTCCGAGGAGAAGGCCAA (SEQ ID NO:174)
Survey_VEGF-fwd	CCACACAGCTTCCCGTTCTC (SEQ ID NO:175)
Survey_VEGF-rev	GAGAGCCGTTCCCTCTTTGC (SEQ ID NO:176)

Primers for High-throughput sequencing of on-target and off-target sites in human genome.

HTS_EMX_ON-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT CCTCCCCATTGGCCTGCTTC (SEQ ID NO:177)
HTS_EMX_Off1-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT TCGTCCTGCTCTCACTTAGAC (SEQ ID NO:178)
HTS_EMX_Off2-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT TTTTGTGGCTTGGCCCCAGT (SEQ ID NO:179)
HTS_EMX_Off3-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT TGCAGTCTCATGACTTGGCCT (SEQ ID NO:180)
HTS_EMX_Off4-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT TTCTGAGGGCTGCTACCTGT (SEQ ID NO:181)
HTS_VEFG_ON-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT ACATGAAGCAACTCCAGTCCCA (SEQ ID NO:182)
HTS_VEGF_Off1-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT AGCAGACCCACTGAGTCAACTG (SEQ ID NO:183)
HTS_VEGF_Off2-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT CCCGCCACAGTCGTGTCAT (SEQ ID NO:184)
HTS_VEGF_Off3-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT CGCCCCGGTACAAGGTGA (SEQ ID NO:185)
HTS_VEGF_Off4-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT GTACCGTACATTGTAGGATGTTT (SEQ ID NO:186)
HTS_CLTA2_ON-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT CCTCATCTCCCTCAAGCAGGC (SEQ ID NO:187)
HTS_CLTA2_Off1-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT ATTCTGCTCTTGAGGTTATTTGT (SEQ ID NO:188)
HTS_CLTA2_Off2-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT CACCTCTGCCTCAAGAGCAGAAAA (SEQ ID NO:189)

HTS_CLTA2_Off3-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT TGTGTGTGTGTGTGTGTAGGACT (SEQ ID NO:190)
HTS_EMX_ON-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT TCATCTGTGCCCCTCCCTCC (SEQ ID NO:191)
HTS_EMX_Off-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT CGAGAAGGAGGTGCAGGAG (SEQ ID NO:192)
HTS_EMX_Off-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT CGGGAGCTGTTCAGAGGCTG (SEQ ID NO:193)
HTS_EMX_Off-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT CTCACCTGGGCGAGAAAGGT (SEQ ID NO:194)
HTS_EMX_Off-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT AAAACTCAAAGAAATGCCCAATCA (SEQ ID NO:195)
HTS_VEFG_ON-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT AGACGCTGCTCGCTCCATTC (SEQ ID NO:196)
HTS_VEGF_Off1-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT ACAGGCATGAATCACTGCACCT (SEQ ID NO:197)
HTS_VEGF_Off2-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT GCGGCAACTTCAGACAACCGA (SEQ ID NO:198)
HTS_VEGF_Off3-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT GACCCAGGGGCACCAGTT (SEQ ID NO:199)
HTS_VEGF_Off4-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT CTGCCTTCATTGCTTAAAAGTGGAT (SEQ ID NO:200)
HTS_CLTA2_ON-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT ACAGTTGAAGGAAGGAAACATGC (SEQ ID NO:201)
HTS_CLTA2_Off1-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT GCTGCATTTGCCCATTTCCA (SEQ ID NO:202)
HTS_CLTA2_Off2-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT GTTGGGGGAGGAGGAGCTTAT (SEQ ID NO:203)
HTS_CLTA2_Off3-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT CTAAGAGCTATAAGGGCAAATGACT (SEQ ID NO:204)
HTS_EGFP-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN ACGTAAACGGCCACAAGTTC (SEQ ID NO:205)
HTS_EGFP-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT GTCGTCCTTGAAGAAGATGGTG (SEQ ID NO:206)
HTS_MusEMX_ON-fwd	CACTCTTTCCCTACACGACGCTCTTCCGATCT CCAGGTGAAGGTGTGGTTCCAG (SEQ ID NO:207)
HTS_MusEMX_ON-rev	GGAGTTCAGACGTGTGCTCTTCCGATCT CCCCTAGTCATTGGAGGTGAC (SEQ ID NO:208)

Results

Highly efficient delivery of Cre recombinase fused to a supernegatively charged protein

[00278] It was speculated that imparting the highly anionic electrostatic properties of nucleic acids to genome-editing proteins may enable their efficient delivery into mammalian cells using cationic lipids (Figure 27(A)). For proteins of interest that are not natively highly negatively charged, it was thought that fusion with a natural or engineered supernegatively

charged protein[17] would impart a polyanionic character. For nucleic acid-binding proteins, it was speculated that simple complexation with native DNA or RNA substrates might provide sufficient anionic character to support cationic lipid-based delivery (Figure 27(A)).

[00279] It was first tested whether the engineered supernegatively charged GFP variant,[35] (–30)GFP, could mediate encapsulation and delivery of fused protein cargo (Figure 27(B)). (–30)GFP was fused to Cre recombinase and several commercially available cationic lipids were tested for their ability to functionally deliver the fusion into HeLa cells that only express DsRed upon Cre-mediated recombination (Figure 28(A)). Delivery of 10 nM (–30)GFP-Cre complexed with 1.5 μL Lipofectamine RNAiMAX (hereafter referred to as "RNAiMAX", Life Technologies, Carlsbad CA) in media containing 10% fetal bovine serum (FBS) led to strong DsRed fluorescence signal among treated cells. Fluorescence-activated cell sorting (FACS) revealed that 48 hours after treatment 52% of cells expressed DsRed consistent with Cre recombination (Figure 28(B)).

[00280] Optimization resulted in recombination efficiencies of 65% using 25 nM (–30)GFP-Cre complexed with 1.5 μL RNAiMAX in 250 μL of media containing 10% FBS (Figure 28(C)). The potency of lipid-mediated anionic Cre delivery is notable compared to that of cationic protein-mediated delivery. Only 1 nM (–30)GFP-Cre with cationic lipid was needed to result in 15-20% recombined cells, whereas 1 μM (+36)GFP-Cre was required to achieve this extent of recombination, representing a 1,000-fold difference in delivery potency (Figure 28(C)). Nearly identical results were observed in a second Cre reporter cell line (BSR TdTomato) (Figure 33(A)). Increasing the amount of cationic lipid increased toxicity (Figure 33(B)) and it was found that 1.5 μL RNAiMAX per 250 μL sample maximized recombination efficiency while inducing minimal cell toxicity. Under these conditions, cationic lipids did not increase the delivery potency of neutral or cationic Cre recombinase fusions (Figure 28(C) and Figure 33(C)), indicating that the strongly negative charge of (–30)GFP-Cre was required to participate in cationic lipid-mediated delivery. It was also observed that increasing the amount of cationic lipid increased the concentration of protein required for maximal recombination, consistent with a model in which deliverable proteins are complexed with specific stiochiometries of cationic lipids (Figure 28(D)). These observations collectively indicate that cationic lipids can mediate the potent delivery of polyanionic proteins into mammalian cells even in the presence of serum.

[00281] To determine if the higher potency of cationic lipid-mediated (–30)GFP-Cre delivery relative to cationic protein-mediated delivery arises from more total protein uptake by cells, or from a higher fraction of functional, non-endosomal protein molecules that enter

cells, flow cytometry was used to measure GFP fluorescence of cells treated with either (+36)GFP-Cre or liposomal (–30)GFP-Cre under their respective optimal Cre delivery conditions. Comparison of cellular fluorescence and recombination efficiency reveals that lipid-mediated functional delivery of (–30)GFP-Cre is 9,800-fold more potent per amount of endocytosed protein than delivery of (+36)GFP-Cre (Figure 34). Taken together, these results suggest that the unusually high potency of lipid-mediated delivery of anionic proteins does not arise from unusually high protein uptake in each cell, but rather from post-endocytosis processes that likely include endosomal escape into the cytoplasm and the avoidance of lysosomal protein degradation.

[00282] To test whether the ability to deliver polyanionic proteins is dependent on proprietary components in RNAiMAX or if other cationic lipids are capable of mediating similarly potent delivery, several other transfection reagents designed to deliver nucleic acids were tested (Figure 28(E)). While RNAiMAX remained the most effective functional delivery agent for (–30)GFP-Cre, other cationic lipid formulations also resulted in potent delivery. Lipofectamine 2000 and Lipofectamine LTX (Life Technologies, Carlsbad CA), two plasmid transfection reagents based on cationic lipid formulations,[21] and SAINT-Red (Synvolux Therapeutics, Groningen Netherlands), an siRNA delivery formulation containing a synthetic pyridium-containing cationic lipid, all resulted in strong functional (–30)GFP-Cre delivery over a range of concentrations (Figure 28(E)). In contrast, strong deliveries with the cationic lipid DOTAP (Roche Diagnostics, Indianapolis IN) or the peptide-based nucleic acid delivery agent EZ-PLEX (Ascension Bio, Tampa FL) were not observed (Figure 28(E)). These observations collectively indicate that several (but not all) cationic lipids are able to encapsulate and deliver negatively charged proteins into human cells.

[00283] It was speculated that it should be possible to use cationic lipids to deliver polyanionic proteins other than (–30)GFP. Engineered polyanionic protein domains commonly used in biomedical research include the VP64 activation domain (–22 net theoretical charge) widely used in fusions with engineered zinc finger arrays, TALE repeat arrays, or dCas9 for transcriptional activation, and 3x FLAG (–7 net theoretical charge), an epitope tag used for protein purification and visualization (Figure 28(F)). It was observed that both VP64 and 3x FLAG enhance functional delivery of Cre recombinase with cationic lipids, though not as effectively as (–30)GFP, likely due to their lower overall negative charge (Figure 33(C)). These observations demonstrate that unusually negatively charged proteins beyond (–30)GFP can mediate efficient cationic lipid-based delivery into mammalian cells.

Functional delivery of TALE activator proteins

[00284] The lipid-mediated delivery of TALE-VP64 transcriptional activators (approximately +4 theoretical net charge, depending on TALE variant used) into cultured human cells was tested. While modestly effective cleavage of endogenous genes by delivered TALEN proteins has been demonstrated in mammalian cells in the absence of serum using cationic peptides such as Arg_9,[36] the delivery of TALE-based transcription factor proteins has not yet been reported, and no effective delivery of TALE proteins in serum has been previously described to our knowledge. The gene for neurotrophin-3 (NTF3), a neural growth factor that has been associated with neurodegenerative diseases, was targeted.[37] A previously described NTF3-targetting TALE-VP64[38] was fused to (–30)GFP (Figure 29(A)) and treated HEK293T cells with 25 nM (–30)GFP-NTF3 TALE1-VP64 and RNAiMAX under the conditions optimized for Cre delivery. Gene expression levels of NTF3 4 hours after treatment were 3.5-fold higher in cells treated with 25 nM (–30)GFP-NTF3 TALE-VP64 and RNAiMAX than untreated cells, cells treated with RNAiMAX only, or cells treated with a VEGF-targeting TALE transcriptional activator (Figure 29(B)). Comparable levels of NTF3 expression were observed 48 hours after transfection of plasmids encoding the same NTF3-targeting TALE-VP64 (Figure 29(B)).

[00285] Since the synergistic expression of multiple TALE activators targeting different sites on the same gene has been shown to augment gene activation,[38] five distinct NTF3-targeting TALE activators fused to (–30) GFP using RNAiMAX were simultaneously delivered. Protein-lipid complexes were prepared as above by adding the five (–30)GFP-NTF3-TALE-VP64 proteins at 5 nM each, for a total of 25 nM protein. A 6.5-fold increase in NTF3 expression was observed after a 4-hour incubation (Figure 29(B) and Figure 35), while plasmid co-transfection of all five NTF3 TALE activators, followed by a 48-hour incubation, resulted in a 10-fold increase in NTF3 expression levels (Figure 29(B)). These findings demonstrate that TALE activator proteins can be delivered using cationic lipids to transiently activate gene expression in human cells. The delivery of programmable transcriptional activator proteins may enable the one-time activation of a target gene while avoiding chronic gene expression, a general concern with DNA-based delivery of programmable transcription factors. This capability may prove especially valuable for proteins that effect a one-time permanent change in cell state or cell fate when transiently expressed.[39]

Highly efficient delivery of Cas9:sgRNA protein:RNA complexes into human cells

[00286] Given the potent lipid-mediated delivery of polyanionic Cre and TALE activator protein variants in full-serum media, it was speculated that CRISPR-Cas9:sgRNA complexes, either as fusions with (–30)GFP or as native polyanionic Cas9:guide RNA complexes, might also be delivered into human cells using this approach. Using a well-established Cas9-induced gene disruption assay,[40] specific sites within a genomic EGFP reporter gene in human U2OS cells were targeted (Figure 36(A)). On-target Cas9 cleavage induces non-homologous end joining (NHEJ) in *EGFP* and the loss of cell fluorescence. To avoid interference from the fluorescence of (–30)GFP, a Y67S mutation was introduced into (–30)GFP to eliminate its fluorescence, and designated this non-fluorescent variant as (–30)dGFP.

[00287] Treatment of U2OS reporter cells with 25 nM (–30)dGFP-NLS-Cas9 and 50 nM EGFP-targeting sgRNA with RNAiMAX in media containing 10% FBS showed loss of EGFP expression in 48% of cells (Figure 30(A)). Cotransfection of plasmids expressing Cas9 or sgRNA resulted in similar EGFP loss in 37% of cells (Figure 30(A)). No significant *EGFP* disruption was observed upon transfection of plasmids encoding EGFP sgRNA alone, Cas9 alone, or cotransfection of plasmids encoding Cas9 and an sgRNA designed to target a *VEGF* locus (Figure 30(A), Figure 36(B)). It was confirmed that the robust disruption of EGFP was not a result of cellular toxicity (Figures 36(C)-(D)). It was also observed that treatment of cells with (+36)dGFP-NLS-Cas9 and sgRNA in the presence of 10% FBS serum did not lead to efficient gene disruption (Figure 30(A)), suggesting that cationic-peptide based methods of delivery for Cas9 and sgRNA are not effective perhaps due to interference of gRNA:Cas9 complex formation or nuclease function by superpositively charged proteins.[41] Together, these results establish that cationic lipid-mediated delivery of (–30)dGFP-NLS-Cas9:sgRNA complexes can result in efficient sgRNA-dependent target gene disruption in human cells.

Polyanionic sgRNA is necessary and sufficient for efficient lipid-mediated Cas9 delivery

[00288] Since the complex of native Cas9 protein (+22 net theoretical charge) and an sgRNA (~103 anionic phosphate groups) should be overall highly anionic, next it was tested if native Cas9:sgRNA complexes without fusion to polyanionic proteins can be delivered into human cells using cationic lipids. Treatment of U2OS EGFP reporter cells with 100 nM Cas9, 100 nM EGFP sgRNA, and 0.8 μL RNAiMAX resulted in 65% disruption of the EGFP reporter gene (Figure 30(A)). Treatment of cells with Cas9 protein and sgRNA, but without RNAiMAX, resulted in no loss of GFP fluorescence (Figure 30(A)). These observations suggest that sgRNA alone, even in the absence of a supernegatively charged fusion protein,

can provide the highly anionic character needed to mediate cationic lipid-based delivery of Cas9.

[00289] Comparison of gene disruption efficiency arising from the cationic lipid-mediated delivery of (–30)dGFP-NLS-Cas9:sgRNA versus Cas9:sgRNA revealed that at low doses (–30)dGFP-NLS-Cas9 results in more efficient gene disruption than native Cas9 (Figure 37(A)), it is outperformed by native Cas9 at higher concentrations, as well as at the respective optimal protein:sgRNA dose of either protein (Figures 37(B)-37(C)). These results further establish that sgRNA can supply sufficient negative charge to support cationic lipid-based delivery of complexed Cas9 protein.

[00290] It was also observed that while overall less protein was required for optimal delivery of (–30)dGFP-NLS-Cas9 than Cas9, a higher sgRNA:protein ratio was required for maximal (–30)dGFP-NLS-Cas9-mediated EGFP gene disruption than for native Cas9-mediated gene disruption (Figure 37(D)). It was speculated that more equivalents of sgRNA are needed to complex with (–30)dGFP-NLS-Cas9 since fused (–30)dGFP may electrostatically interfere with Cas9:sgRNA complexation. As the ideal protein dose for (–30)dGFP-NLS-Cas9 mediated EGFP gene disruption is 10-fold lower than that of wild-type Cas9, the results also suggest that (–30)dGFP-Cas9 is better encapsulated by cationic liposomes than Cas9:sgRNA due to its higher overall negative charge, but this charge magnitude may interfere with Cas9:sgRNA interactions, necessitating more sgRNA per protein and potentially reducing total delivered Cas9 activity. In addition, NLS-Cas9 and Cas9-NLS proteins were generated and tested, and it was observed that while the presence of an NLS in (–30)dGFP-NLS-Cas9 could at least partially explain differences in delivery efficacy at very low concentrations, Cas9, NLS-Cas9, and Cas9-NLS all result in higher efficiency of *EGFP* disruption than (–30)dGFP-NLS-Cas9 at 25 nM or higher concentrations (Figures 38(A)-(C)).

[00291] Cas9:sgRNA delivery with cationic lipid formulations other than RNAiMAX was also tested. Delivery with Lipofectamine 2000 was notably more efficient than with RNAiMAX, resulting in up to 80% Cas9-mediated gene disruption (Figure 39(A)), and maintaining high efficiency (60% gene disruption) even at 1 nM protein (Figure 39(A)). However, due to the somewhat higher toxicity of Lipofectamine 2000 compared to RNAiMAX under cell culture conditions (Figures 33(B)-C)), RNAiMAX was used for all subsequent cell culture studies.

[00292] To verify that EGFP disruption arose from genome modification and not only from Cas9 binding,[42] the T7 endonuclease I (T7EI) assay[43] was used to detect and quantify

the frequency of Cas9-mediated genomic insertion/deletion mutations (indels) at the target EGFP locus (Figure 30(B)). The T7EI assay results showed that only those cells treated with both Cas9 and EGFP sgRNA plasmids, or Cas9 protein and purified EGFP sgRNA, contained indels at the target site. Taken together, these findings establish that active Cas9:sgRNA complexes can be potently delivered into human cells with cationic lipids in a manner dependent on the negative charge provided by the sgRNA.

[00293] U2OS EGFP reporter cells were also treated with a single lipid-mediated delivery treatment of Cas9 complexed with a mixture of four gRNAs targeting *EGFP*, *CLTA*, *EMX*, and *VEGF*. This treatment resulted in efficient disruption of all four targets, with cleavage efficiencies of 58%, 28%, 16%, and 40%, respectively, as measured by T7E1 cleavage assay. These high gene disruption efficiencies from a single delivery of 50 nM Cas9 and 25 nM of each sgRNA (100 nM total sgRNA) demonstrate that lipid-mediated Cas9:sgRNA delivery can support efficient multiplexed genome editing (Figure 30(C)).

Functional delivery of Cas9 nickases and dCas9 activators

[00294] Next,whether cationic lipid-based protein delivery could be extended to deliver other Cas9-derived genome engineering tools such as Cas9 nickases[44] and Cas9-based transcriptional activators was tested.[45] Gene disruption efficiency in U2OS EGFP reporter cells resulting from delivery of Cas9 D10A nickase was measured, either by cotransfection of nickase and appropriate paired EGFP-targeting sgRNA plasmids, or as purified protein complexed with pairs of EGFP sgRNAs using RNAiMAX (Figure 30(D)). Both plasmid and cationic lipid-mediated protein:RNA delivery of dual Cas9 nickases resulted in EGFP disruption with similar efficiencies (Figure 30(D)) only in the presence of sgRNA pairs targeting opposite strands, (sgRNA pairs g1+g5, and g3+g7), but not with sgRNA pairs targeting the same strand (sgRNA pair g5+g7) (Figure 30(D)), consistent with previous reports of Cas9 nickase cleavage requirements.[46]

[00295] The *NTF3* transcriptional activation efficiencies in HEK293T cells resulting from either plasmid transfection or direct protein:sgRNA complex delivery of dCas9 fused to a VP64 activation domain were also compared.[45] Delivery of dCas9-VP64 activators either by plasmid transfection or RNAiMAX-mediated protein delivery resulted in strong (≥ ~10-fold) activation of *NTF3* transcription (Figure 30(E) and Figure 40). Transcriptional activation levels resulting from plasmid transfection were more potent than activation resulting from protein delivery at optimal assay times for each delivery method (Figure 30(E)), potentially due to the sustained expression both Cas9 activator protein and sgRNA

from the plasmids compared to the transient, single dose of purified protein and RNA. While the above results indicate that such factors do not limit the potency of irreversible genome modification by delivered Cas9 nuclease and nickase proteins (Figures 40(A) and 40(D)), the low dose and transient nature of the delivered protein may more strongly limit potency of dynamic processes such as transcriptional activation. Nevertheless, these results collectively indicate that both Cas9 nickases and Cas9 transcriptional activators can also be delivered effectively by cationic lipid-mediated protein:RNA complex delivery.

Cas9:sgRNA delivery modifies genomes with greater specificity than DNA transfection

[00296] DNA-free delivery of functional Cas9:sgRNA complexes circumvents risks associated with viral or other gene delivery methods and has the potential to improve the specificity of genome modification by avoiding the unnecessary expression of genome-editing agent after the target locus is modified. To test if the described approach can disrupt endogenous genes in human cells, genomic loci in the *EMX1, CLTA2,* and *VEGF* genes were targeted due to their potential biomedical relevance and their use in previous studies[40,46,47] of Cas9 off-target cleavage activity. Cationic lipid-mediated delivery of Cas9:sgRNA complexes into HEK293T cells resulted in robust cleavage of all three human genes with efficiencies comparable to or greater than those of plasmid transfection methods as revealed by the T7EI assay using the same Cas9:sgRNA delivery conditions previously optimized for U2OS cells (Figure 31(A)).

[00297] To compare the endogenous gene modification specificity of plasmid versus protein:RNA delivery methods for Cas9, the on-target locus was amplified as well as several known off-target sites (Figure 44) from genomic DNA isolated from HEK293 cells treated either by transfection of Cas9 and sgRNA expression plasmids, or by RNAiMAX-mediated Cas9:sgRNA complex delivery under conditions that resulted in comparable on-target modification efficiencies. The indel frequencies at the three on-target and 11 off-target sites were assayed by high-throughput DNA sequencing (Figure 45). For all three target genes, the frequency of on-target DNA modification resulting from either plasmid or protein:sgRNA delivery was approximately 10% (Figures 41(A)-(C)), enabling a comparison of off-target modification between the two techniques under treatment conditions that result in very similar on-target genome modification efficiencies. Importantly, the frequency of off-target genome modification for all 11 off-target sites was lower from protein:sgRNA delivery compared with plasmid delivery, and as a result the ratio of on-target to off-target

modification for all sites tested was up to 19-fold higher for protein:sgRNA delivery than for plasmid delivery (Figures 31(B)-(D)).

[00298] DNA modification specificity was higher for protein:sgRNA delivery than for plasmid delivery at loci with high levels of off-target modification (such as the four *VEGF* off-target sites, for which plasmid delivery yielded average on-target:off-target modification ratios between 4- and 20-fold but protein:RNA delivery yielded average on-target:off-target modification ratios between 9- and 400-fold) as well as for loci with lower levels of off-target modification (such as the three *EMX* off-target loci, for which plasmid delivery yielded average on-target:off-target modification ratios as low as 64-fold but protein:RNA delivery yielded average on-target:off-target modification ratios of 500- to 2,000-fold). Taken together, these results indicate that the delivery of Cas9:sgRNA complexes using cationic lipids can effect target gene modification at high efficiency and with substantially greater specificity than the delivery of DNA expressing Cas9 and sgRNA.

Delivery of Cas9:sgRNA into mouse embryonic stem cells

[00299] The potent and transient cationic lipid-mediated delivery of Cas9:sgRNA to effect efficient, permanent, and highly specific gene editing could be especially useful in stem cells. To test this possibility, mouse embryonic stem cells expressing Tau-EGFP[48] were treated with Cas9 and an *EGFP*-targeting sgRNA. Under standard stem-cell culture conditions, EGFP-positive floating spheres were formed. The floating spheres were treated with Cas9:sgRNA complexed with RNAiMAX, or with Cas9 and RNAiMAX without sgRNA as a control. Three days post-treatment, a reduction in GFP fluorescence in the Cas9:sgRNA-treated spheres compared to the control samples was observed (Figure 42(A)). The treated spheres were dissociated, and the cells were allowed to attach to a gelatin-coated dish and differentiate into progenitor cells. Immunohistochemistry using an anti-GFP antibody confirmed knockdown of EGFP expression in the cells of Cas9:sgRNA treated samples, with many nuclei lacking any apparent EGFP. In contrast, all cells derived from control spheres were EGFP positive (Figure 42(B)). Genomic DNA harvested from Cas9:sgRNA-treated cells was subjected to T7EI assay, resulting in clear evidence of indels at the Tau-EGFP locus (Figure 42(C)). From this assay, an indel frequency of 42% was calculated from both cationic lipid-mediated Cas9:sgRNA delivery and transfection of Cas9 and sgRNA DNA. No target modification was detected in control samples lacking Cas9:sgRNA or containing Cas9 and an unrelated gRNA. These findings demonstrate that

cationic lipid-mediated Cas9:sgRNA delivery can effect highly efficient gene disruption in mouse embryonic stem cells.

In vivo cationic lipid-mediated delivery of Cre recombinase and Cas9:sgRNA

[00300] The high-efficiency delivery of functional genome-editing proteins *in vivo* enables a wide range of applications including non-viral therapeutic genome editing to correct genetic diseases. To evaluate the protein delivery method described above in a living mammal, delivery to the mouse inner ear was chosen, due to its confined space, well-characterized inner ear cell types, and the existence of genetic deafness mouse models that may enable future hearing recovery studies. The *in vivo* deliveries of two types of proteins into the mouse inner year were attempted. First, the delivery of (–30)GFP-Cre protein was tested to assess the targeting of inner ear cell types and the efficiency of functional protein delivery. Second, the delivery of Cas9:sgRNA complexes to the inner ear were evaluated to determine if cationic lipid-mediated protein:gRNA complex delivery can support CRISPR-based gene editing *in vivo*.

[00301] It has been previously shown that (+36)GFP-Cre can be delivered to mouse retina,[16] although the protein resulted in only modest levels of recombinant conversion suggestive of inefficient *in vivo* delivery. For our initial inner ear delivery trials, (–30)GFP-Cre was complexed with RNAiMAX and the complex was injected into the cochlea of postnatal day 0 (P0) reporter mice with a genomically integrated floxed-STOP tdTomato reporter. As with the previously described *in vitro* Cre reporter cell line, functional delivery of Cre to the inner ear cells, followed by endosomal escape, nuclear localization, and Cre-mediated recombination results in expression of tdTomato. After injection, the cochleas were harvested for immunolabeling with inner ear cell markers for co-localization with tdTomato. RNAiMAX injection alone was used as control. Five days following injection of (–30)GFP-Cre and RNAiMAX, cochlear outer hair cells, the auditory sensory cells that detect sound, showed strong tdTomato signal that co-localized with the hair cell marker myosin VIIa (Myo7a), demonstrating functional Cre delivery to hair cells (Figures 32(A)-(B)). No tdTomato expression was detected in control cochleas (Figure 32(A)). The tdTomato signal was concentrated in the region of the injection site at the basal turn of the cochlea. On average 33±3% of outer hair cells were tdTomato positive at the base of the cochlea ($P < 0.001$; mean ± SEM, $n = 4$).

[00302] To further determine the effect of cationic lipid-mediated (–30)GFP-Cre protein delivery on targeted cells, hair cell stereocilia, a delicate structure that is essential for

hearing, were examined 10 days post-injection. TdTomato positive outer hair cells had typical stereocilia structure as imaged by espin expression, similar to control stereocilia (Figure 32(B)). No tdTomato expression was detected in control cochleas. These observations indicate that cationic lipid-mediated delivery of (–30)GFP-Cre protein effects recombination in cochlear outer hair cells without apparently affecting hair cell architecture.

[00303] Because target volume, protein dose, and sgRNA dose *in vivo* are different than in cell culture experiments, the above experiments were repeated under different delivery conditions. Delivery using Lipofectamine 2000 was tested due to its higher potency *in vitro* (Figure 39(A)) and dramatically higher recombination efficiency was observed: over 90% outer hair cells in cochleas treated with (–30)GFP-Cre + Lipofectamine 2000 were tdTomato positive (Figure 32(C)). In comparison to control samples, some outer hair cell loss was observed (Figure 32(C)), consistent with the previous observation of the higher cell toxicity of Lipofectamine 2000, although the overall cochlear architecture was preserved.

[00304] To test the effectiveness of Cas9:sgRNA delivery *in vivo*, Cas9 and sgRNA targeting *EGFP* were combined with RNAiMAX and the resulting complexes were injected into postnatal day 2 (P2) transgenic Atoh1-GFP mouse cochlea in which all hair cells express GFP under the control of a hair cell-specific enhancer for transcription factor Atoh1.[49] Using this model, Cas9:sgRNA-mediated disruption of *EGFP* results in loss of EGFP fluorescence in outer hair cells. Ten days after injection of Cas9:sgRNA with cationic lipid, the absence of GFP was observed in 13% of outer hair cells near the injection site. In contrast, control cochlea injected with Cas9 protein and RNAiMAX without any sgRNA showed no loss of EGFP signal (Figure 32(D)). The outer hair cells of cochlea injected with Cas9:sgRNA RNAiMAX complexes appeared to be otherwise unaffected, with stereotypical expression of Myo7a and healthy nuclei, consistent with minimal hair cell toxicity (Figure 32(D)). High-throughput DNA sequencing of genomic DNA isolated from cochlea tissue samples revealed indels consistent with *GFP* target gene disruption in the treated samples, but not in the control samples that lacked sgRNA (Figure 43(A)). In addition, the inner ear *in vivo* delivery of Cas9:sgRNA using an sgRNA that targets the *EMX* gene was repeated and indels in the *EMX* gene in treated animals, but not control animals were similarly observed (Figure 43(B))

[00305] As (–30)GFP-Cre complexed with Lipofectamine 2000 resulted in more efficient modification of the target hair cell population than (–30)GFP-Cre complexed with RNAiMAX (Figures 32(A) and 32(C)), its use on Cas9:sgRNA delivery to Atoh1-GFP cochlea was tested as above. Loss of GFP expression was observed in 20% of outer hair cells

near the injection site after 10 days, whereas all outer hair cells maintained strong GFP expression in control cochlea injected with Cas9 and Lipofectamine 2000 but no sgRNA (Figure 32(D)). In contrast to modest hair cell loss observed following Lipofectamine 2000 delivery of (–30)GFP-Cre (Figure 32(C)), outer hair cells targeted by Cas9:sgRNA exhibited no obvious toxicity or structural alteration (Figure 32(D)).

[00306] As with (–30)GFP-Cre, virus-free, cationic lipid-mediated delivery of Cas9:sgRNA into the mouse inner ear successfully modified a specific genomic locus in the outer hair cell population, leading to loss of target gene expression. Nearly half of all types of genetic deafness arise from hair cell loss or dysfunction,[50] the results presented herein suggest a potential strategy based on the delivery of Cas9:sgRNA complexes to genetically modify these cells to effect hearing recovery.

Determination of protein delivery efficacy for (–30)GFP-Cre

[00307] To determine if the higher potency of liposome-mediated (–30)GFP-Cre delivery compared with that of cationic protein delivery arises from more total protein uptake by cells or from a higher fraction of functional, non-endosomal protein molecules taken up by the cells, flow cytometry was used to measure GFP fluorescence of cells treated with either (+36)GFP-Cre or liposomal (–30)GFP-Cre under their respective optimal Cre delivery conditions. Cell fluorescence reports total endocytosed (–30)GFP-Cre or (+36)GFP-Cre regardless of endosomal or non-endosomal localization.[1] Lipid-mediated protein delivery resulted in surprisingly small increases in total protein uptake (Figure 34(A)), despite the high efficiency of lipid-mediated functional Cre delivery. While (+36)GFP-Cre treatment increased cellular GFP fluorescence by up to three orders of magnitude in a dose-dependent manner (Figure 34(A)), consistent with previous reports,[1,2] liposomal (–30)GFP-Cre treatment induced at most 5-fold increases in cellular GFP fluorescence (Figure 34(A)). Comparison of cellular fluorescence and recombination efficiency reveals that lipid-mediated functional delivery of (–30)GFP-Cre is 9,800-fold more potent per amount of endocytosed protein than delivery of (+36)GFP-Cre (Figure 34(B)).

[00308] To test if complexation of anionic (–30)GFP with cationic lipids interferes with GFP fluorescence and thus masks the true amount of cargo that enters the cell mCherry, which is fluorescent but not highly anionic, was fused to either (–30)GFP or (+36)GFP and delivered both protein fusions to HeLa cells. After washing away protein that may have adhered to cell surface but did not enter the cell with PBS + heparin (20 U/mL), the cells were analyzed by FACS for mCherry fluorescence 4 hours and 24 hours after treatment. It

was observed that lipid-mediated delivery of (–30)GFP-fused mCherry results in only slight increases in cellular mCherry fluorescence, whereas mCherry fluorescence upon delivery of (+36)GFP-mCherry was generally ≥ 100-fold higher (Figure 34(C)) suggesting that fusion to (–30)GFP does not cause substantial amounts of protein cargo to enter the cell. Moreover, addition of lipids to (–30)GFP-Cre did not measurably alter the GFP fluorescence signal (Figure 34(D)), despite the fact that cationic lipids and anionic (–30)GFP clearly interact. Taken together, these results suggest that the unusually high potency of lipid-mediated delivery of anionic proteins does not arise from unusually high protein uptake in each cell, but rather from post-endocytosis processes that likely include avoidance of protein degradation and endosomal escape into the cytoplasm.

Sensitivity limit of off-target cleavage assays

[00309] The sensitivity of the high-throughput sequencing method for detecting genomic off-target cleavage is limited by the amount genomic DNA (gDNA) input into the PCR amplification of each genomic target site. A 1 ng sample of human gDNA represents only approximately 330 unique genomes, and thus only approximately 330 unique copies of each genomic site are present. PCR amplification for each genomic target was performed on a total of 150 ng of input gDNA, which provides amplicons derived from at most 50,000, unique gDNA copies, respectively. Therefore, the high-throughput sequencing assay cannot detect rare genome modification events that occur at a frequency of less than 1 in 50,000 (0.002%). This limit is noted in Table 2.

[00310] Taken together, these findings suggest that cationic lipid-mediated delivery of genome-editing proteins can serve as a powerful tool and an *in vivo* strategy for the treatment of genetic disease.

Conclusions

[00311] Efficient intracellular protein delivery *in vitro* and especially *in vivo* has been a persistent challenge in biomedical research and protein therapeutics. While delivery using cationic peptides and proteins has been widely studied for over two decades, sensitivity to serum proteins, neutralization by antibodies, degradation by extracellular and intracellular proteases, and poor endosomal escape post-internalization have limited the scope of protein delivery applications using that approach.

[00312] In the current Example, a general strategy for protein delivery that makes use of anionic protein complexation with cationic liposomes is demonstrated. This method was used to deliver diverse protein classes, including the Cre tyrosine recombinase, TALE transcriptional activators, and Cas9 nucleases, nickases, and transcriptional activators (Figure 27(A)) to cultured cell lines, stem cell colonies, and therapeutically relevant *in vivo* sites within the mouse inner ear. The described approach is highly efficient, producing modification rates on par with established nucleic acid transfection methods in cell culture, and enabling Cre recombinase and Cas9-mediated genome modification rates of up to 90% and 20%, respectively, within the inner ear hair cell population of live mice (Figures 32(C)-(D)). These results also suggest that it may be possible to use cationic lipids to efficiently deliver other nucleic acid-binding proteins, including transcription factors that induce therapeutically relevant changes in cell fate, by complexing them with nucleic acids.

[00313] Cationic lipid-based anionic protein delivery outperforms a potent cationic protein delivery fusion partner, (+36)GFP, by up to 9,800-fold per amount of endocytosed protein, inducing more efficient modification of treated cells with orders of magnitude lower doses of protein (Figures 28(C) 34). For Cas9 nuclease delivery, this approach also results in >10-fold more specific genome modification than traditional plasmid transfection (Figures 31(B)-(D)), likely due to the transient window of Cas9 activity to which each genome is exposed compared to DNA delivery methods, consistent with previous reports.[51]

[00314] The described approach is simple to implement, requiring only the purified deliverable protein and the use of popular commercial nucleic acid transfection reagents (Figure 27(B)). Rendering a given protein amenable to this approach requires simple translational fusion to a highly anionic partner, such as (–30)GFP (Figure 27(A)), and is even effective with common translational fusion tags including the VP64 activation domain, and the 3xFLAG affinity tag (Figure 28(F) and Figure 33(C)). In certain cases, as with the Cas9 protein, pre-complexation with a cognate nucleic acid (sgRNA in this case) is sufficient (Figure 30(A)), as the partially exposed bound nucleic acid likely provides sufficient anionic charge to mediate complexation with cationic lipids.

[00315] Others groups have reported the *in vivo* delivery of Cas9 expression constructs in DNA or mRNA form.[52,53] The present Example demonstrates that protein delivery is a viable approach to *in vivo* genome editing.

Table 1.

EMX_On	GAGTCCGAGCAGAAGAAGAAGGG (SEQ ID NO:209)
EMX_Off1	GAG**g**CCGAGCAGAAGAA**ag**ACGG (SEQ ID NO:210)
EMX_Off2	GAGTCC**t**AGCAG**g**AGAAGAAG**a**G (SEQ ID NO:211)
EMX_Off3	GAGTC**ta**AGCAGAAGAAGAAG**a**G (SEQ ID NO:212)
EMX_Off4	GAGT**ta**GAGCAGAAGAAGAAAGG (SEQ ID NO:213)
VEGF_On	GGGTGGGGGGAGTTTGCTCCTGG (SEQ ID NO:214)
VEGF_Off1	GG**a**TGG**a**GGGAGTTTGCTCCTGG (SEQ ID NO:215)
VEGF_Off2	GGG**a**GGG**t**GGAGTTTGCTCCTGG (SEQ ID NO:216)
VEGF_Off3	**c**GG**g**GG**a**GGGAGTTTGCTCCTGG (SEQ ID NO:217)
VEGF_Off4	GGG**ga**GGGG**a**AGTTTGCTCCTGG (SEQ ID NO:218)
CLTA_On	GCAGATGTAGTGTTTCCACAGGG (SEQ ID NO:219)
CLTA_Off1	**a**CA**a**ATGTAGT**a**TTTCCACAGGG (SEQ ID NO:220)
CLTA_Off2	**c**CAGATGTAGT**a**TT**c**CCACAGGG (SEQ ID NO:221)
CLTA_Off3	**ct**AGATG**a**AGTG**c**TTCCACATGG (SEQ ID NO:222)

Table 1. On-target and known off-target substrates of Cas9:sgRNAs that target sites in EMX, VEGF, and CLTA. A list of genomic on-target and off-targets sites of the EMX, VEGF, and CLTA are shown with mutations from the on-target sequence shown in lower case and bold. PAMs are shown in underline.

Table 2.

CLTA Sites	**Mock treatment**	**Plasmid transfection**	**Protein:sgRNA delivery**
CLTA_On Indels	14	1228	1498

Total	10000	10000	10000
Modified (%)	0.140	12.280	14.980
P-value		<1.0E-300	<1.0E-300
On:off specificity	1	1	1
CLTA_Off1			
Indels	7	29	14
Total	41518	205204	125370
Modified (%)	0.017	0.014	0.011
P-value		6.6E-01	4.5E-01
On:off specificity		869	1341
CLTA_Off2			
Indels	5	11	8
Total	25338	83944	54409
Modified (%)	0.020	0.013	0.015
P-value		5.5E-01	5.7E-01
On:off specificity		937	1019
CLTA_Off3			
Indels	6	22	8
Total	41643	189886	76863
Modified (%)	0.014	0.012	0.010
P-value		6.2E-01	5.8E-01
On:off specificity		1060	1439
EMX Sites	**Mock treatment**	**Plasmid transfection**	**Protein:sgRNA delivery**
EMX_On			
Indels	3	930	1140
Total	10000	10000	10000
Modified (%)		0.030	9.300
P-value		1.6E-264	<1.0E-300
On:off specificity	1	1	1
EMX_Off1			
Indels	0	6	6
Total	24623	90935	100778
Modified (%)		<0.002	0.007
P-value		3.5E-01	6.1E-01
On:off specificity		1409	1915
EMX_Off2			
Indels	16	53	38
Total	36061	204068	130084
Modified (%)		0.044	0.026
P-value		6.4E-02	1.8E-01
On:off specificity		358	390
EMX_Off3			
Indels	20	147	44
Total	32575	157848	110878
Modified (%)		0.061	0.093
P-value		8.1E-02	1.3E-01
On:off specificity		100	287
EMX_Off4			
Indels	16	141	23
Total	45548	86586	73451
Modified (%)		0.035	0.163

P-value		2.8E-12	7.4E-01
On:off specificity		57	364
VEGF Sites	**Mock treatment**	**Plasmid transfection**	**Protein:sgRNA delivery**
VEGF_On			
Indels	1	989	785
Total		10000	10000
Modified (%)	0.010	9.890	7.850
P-value		1.5E-285	5.7E-228
On:off specificity	1	1	1
VEGF_Off1			
Indels	4	4240	602
Total		38625	184554
Modified (%)	0.010	2.297	0.394
P-value		<1.0E-300	3.7E-52
On:off specificity		4	20
VEGF_Off2			
Indels	5	727	18
Total		30301	79164
Modified (%)	0.017	0.918	<0.002
P-value		4.7E-93	1.3E-04
On:off specificity		11	3925
VEGF_Off3			
Indels	2	536	21
Total		26379	110902
Modified (%)	0.008	0.483	0.022
P-value		2.0E-46	2.0E-01
On:off specificity		20	352
VEGF_Off4			
Indels	0	1531	45
Total		26012	122403

Table 2. Indel frequencies, P values, and on-target:off-target cleavage specificity ratios for *EMX*, *CLTA*, and *VEGF* on-target sites and 11 known off-target sites. CLTA sites: Total: total number of sequence counts; only the first 10,000 sequences were analyzed for the on-target site sequences. Modified: number of indels divided by total number of sequences as percentages. Upper limits of potential modification were calculated for sites with no observed indels by assuming there is less than one indel then dividing by the total sequence count to arrive at an upper limit modification percentage, or taking the theoretical limit of detection (1/49,500; see **Results** above), whichever value was larger. P-values: for mock treatment, Cas9 plasmid transfection, and liposomal Cas9 protein:sgRNA delivery, P-values were calculated as using a two-sided Fisher's exact test between each *CLTA*-targeted treatment sample (either DNA transfection or protein:sgRNA delivery) versus the control sample (mock treatment) treated with Cas9 protein and an sgRNA targeting *EGFP*. On:off specificity is the ratio of on-target to off-target genomic modification frequency for each site. EMX sites

shows the experimental and analytic methods of CLTA analysis applied to *EMX* target sites. VEGF sites shows the experimental and analytic methods of CLTA analysis as applied to *VEGF* target sites. Indel numbers in the mock treatment control were subtracted from both plasmid transfection and protein:sgRNA delivery indel numbers for determining total number of indels and for calculating on-target:off-target ratios in Figure 31 in the main text and also for Figure 41.

References

1. Putney, S. D. & Burke, P. A. Improving protein therapeutics with sustained-release formulations. *Nat. Biotechnol.* **16,** 153–157 (1998).

2. Mullen, L. *et al.* Latent cytokines for targeted therapy of inflammatory disorders. *Expert Opin. Drug Deliv.* **11,** 101–110 (2014).

3. Song, E. *et al.* Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. *Nat. Biotechnol.* **23,** 709–717 (2005).

4. Leader, B., Baca, Q. J. & Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discov.* **7,** 21–39 (2008).

5. Hartung, S. D. *et al.* Correction of Metabolic, Craniofacial, and Neurologic Abnormalities in MPS I Mice Treated at Birth with Adeno-associated Virus Vector Transducing the Human α-L-Iduronidase Gene. *Mol. Ther.* **9,** 866–875 (2004).

6. Wang, J. *et al.* Neutralizing antibodies to therapeutic enzymes: considerations for testing, prevention and treatment. *Nat. Biotechnol.* **26,** 901–908 (2008).

7. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11,** 636–646 (2010).

8. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–355 (2014).

9. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31,** 397–405 (2013).

10. Midoux, P., Pichon, C., Yaouanc, J.-J. & Jaffrès, P.-A. Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers. *Br. J. Pharmacol.* **157,** 166–178 (2009).

11. Bodles-Brakhop, A. M., Heller, R. & Draghia-Akli, R. Electroporation for the Delivery of DNA-based Vaccines and Immunotherapeutics: Current Clinical Developments. *Mol. Ther.* **17,** 585–592 (2009).

12. Kay, M. A., Glorioso, J. C. & Naldini, L. Viral vectors for gene therapy: the art of

turning infectious agents into vehicles of therapeutics. *Nat. Med.* **7,** 33–40 (2001).

13. Zangi, L. *et al.* Modified mRNA directs the fate of heart progenitor cells and induces vascular regeneration after myocardial infarction. *Nat. Biotechnol.* **31,** 898–907 (2013).

14. Wadia, J. S., Stan, R. V. & Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **10,** 310–315 (2004).

15. Daniels, D. S. & Schepartz, A. Intrinsically cell-permeable miniature proteins based on a minimal cationic PPII motif. *J. Am. Chem. Soc.* **129,** 14578–14579 (2007).

16. Cronican, J. J. *et al.* Potent delivery of functional proteins into Mammalian cells in vitro and in vivo using a supercharged protein. *ACS Chem. Biol.* **5,** 747–752 (2010).

17. Thompson, D. B., Cronican, J. J. & Liu, D. R. Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. *Methods Enzymol.* **503,** 293–319 (2012).

18. Thompson, D. B., Villaseñor, R., Dorr, B. M., Zerial, M. & Liu, D. R. Cellular uptake mechanisms and endosomal trafficking of supercharged proteins. *Chem. Biol.* **19,** 831–843 (2012).

19. Heitz, F., Morris, M. C. & Divita, G. Twenty years of cell-penetrating peptides: from molecular mechanisms to therapeutics. *Br. J. Pharmacol.* **157,** 195–206 (2009).

20. Caron, N. J. *et al.* Intracellular delivery of a Tat-eGFP fusion protein into muscle cells. *Mol. Ther. J. Am. Soc. Gene Ther.* **3,** 310–318 (2001).

21. Chesnoy, S. & Huang, L. Structure and function of lipid-DNA complexes for gene delivery. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 27–47 (2000).

22. Al-Taei, S. *et al.* Intracellular traffic and fate of protein transduction domains HIV-1 TAT peptide and octaarginine. Implications for their utilization as drug delivery vectors. *Bioconjug. Chem.* **17,** 90–100 (2006).

23. Shete, H. K., Prabhu, R. H. & Patravale, V. B. Endosomal escape: a bottleneck in intracellular delivery. *J. Nanosci. Nanotechnol.* **14,** 460–474 (2014).

24. Aguilera, T. A., Olson, E. S., Timmers, M. M., Jiang, T. & Tsien, R. Y. Systemic in vivo distribution of activatable cell penetrating peptides is superior to that of cell penetrating peptides. *Integr. Biol. Quant. Biosci. Nano Macro* **1,** 371–381 (2009).

25. Coelho, T. *et al.* Safety and efficacy of RNAi therapy for transthyretin amyloidosis. *N. Engl. J. Med.* **369,** 819–829 (2013).

26. Judge, A. D., Bola, G., Lee, A. C. H. & MacLachlan, I. Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. *Mol. Ther. J. Am. Soc. Gene Ther.*

13, 494–505 (2006).

27. Basha, G. *et al.* Influence of cationic lipid composition on gene silencing properties of lipid nanoparticle formulations of siRNA in antigen-presenting cells. *Mol. Ther. J. Am. Soc. Gene Ther.* **19,** 2186–2200 (2011).

28. Semple, S. C. *et al.* Rational design of cationic lipids for siRNA delivery. *Nat. Biotechnol.* **28,** 172–176 (2010).

29. Boeckle, S., Fahrmeir, J., Roedl, W., Ogris, M. & Wagner, E. Melittin analogs with high lytic activity at endosomal pH enhance transfection with purified targeted PEI polyplexes. *J. Control. Release Off. J. Control. Release Soc.* **112,** 240–248 (2006).

30. Allen, T. M. & Cullis, P. R. Liposomal drug delivery systems: from concept to clinical applications. *Adv. Drug Deliv. Rev.* **65,** 36–48 (2013).

31. Zelphati, O. *et al.* Intracellular delivery of proteins with a new lipid-mediated delivery system. *J. Biol. Chem.* **276,** 35103–35110 (2001).

32. Adrian, J. E. *et al.* Targeted SAINT-O-Somes for improved intracellular delivery of siRNA and cytotoxic drugs into endothelial cells. *J. Control. Release Off. J. Control. Release Soc.* **144,** 341–349 (2010).

33. Morris, M. C., Depollier, J., Mery, J., Heitz, F. & Divita, G. A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat. Biotechnol.* **19,** 1173–1176 (2001).

34. Colletier, J.-P., Chaize, B., Winterhalter, M. & Fournier, D. Protein encapsulation in liposomes: efficiency depends on interactions between protein and phospholipid bilayer. *BMC Biotechnol.* **2,** 9 (2002).

35. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129,** 10110–10112 (2007).

36. Liu, J., Gaj, T., Patterson, J. T., Sirk, S. J. & Barbas III, C. F. Cell-Penetrating Peptide-Mediated Delivery of TALEN Proteins via Bioconjugation for Genome Engineering. *PLoS ONE* **9,** e85755 (2014).

37. Tessarollo, L., Vogel, K. S., Palko, M. E., Reid, S. W. & Parada, L. F. Targeted mutation in the neurotrophin-3 gene results in loss of muscle sensory neurons. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11844–11848 (1994).

38. Maeder, M. L. *et al.* Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods* **10,** 243–245 (2013).

39. Jopling, C., Boue, S. & Belmonte, J. C. I. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nat. Rev. Mol. Cell Biol.* **12,** 79–89 (2011).

40. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32,** 279–284 (2014).

41. McNaughton, B. R., Cronican, J. J., Thompson, D. B. & Liu, D. R. Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 6111–6116 (2009).

42. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).

43. Guschin, D. Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol. Clifton NJ* **649,** 247–256 (2010).

44. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154,** 1380–1389 (2013).

45. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).

46. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32,** 577–582 (2014).

47. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31,** 839–843 (2013).

48. Li, H. *et al.* Differentiation of neurons from neural precursors generated in floating spheres from embryonic stem cells. *BMC Neurosci.* **10,** 122 (2009).

49. Lumpkin, E. A. *et al.* Math1-driven GFP expression in the developing nervous system of transgenic mice. *Gene Expr. Patterns GEP* **3,** 389–395 (2003).

50. Van Camp, G. & Smith, R. Hereditary Hearing Loss. at <http://hereditaryhearingloss.org>

51. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24,** 1012–1019 (2014).

52. Yin, H. *et al.* Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* **32,** 551–553 (2014).

53. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153,** 910–918 (2013).

54. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).

55. Sage, C. *et al.* Proliferation of functional hair cells in vivo in the absence of the

retinoblastoma protein. *Science* **307,** 1114–1118 (2005).

56. Sander, J. D. *et al.* In silico abstraction of zinc finger nuclease cleavage profiles reveals an expanded landscape of off-target sites. *Nucleic Acids Res.* **41,** e181 (2013).

EQUIVALENTS AND SCOPE

[00316] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments, described herein. The scope of the present invention is not intended to be limited to the above Description, but rather is as set forth in the appended claims.

[00317] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments in accordance with the invention described herein. The scope of the present invention is not intended to be limited to the above Description, but rather is as set forth in the appended claims.

[00318] In the claims articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process. Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, *etc.*, from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for any of the purposes disclosed herein are included, and methods of making the composition according to any of the methods of making disclosed herein or other methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[00319] Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be

removed from the group. It should it be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, *etc.*, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, *etc.* For purposes of simplicity those embodiments have not been specifically set forth *in haec verba* herein. It is also noted that the term "comprising" is intended to be open and permits the inclusion of additional elements or steps.

[00320] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or subrange within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

[00321] In addition, it is to be understood that any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the claims. Since such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of the compositions of the invention (*e.g.*, any supercharged protein; any nucleic acid; any method of production; any method of use; *etc.*) can be excluded from any one or more claims, for any reason, whether or not related to the existence of prior art.

[00322] All cited sources, for example, references, publications, databases, database entries, and art cited herein, are incorporated into this application by reference, even if not expressly stated in the citation. In case of conflicting statements of a cited source and the instant application, the statement in the instant application shall control.

CLAIMS

What is claimed is:

1. A composition comprising a supercharged protein associated with an RNA-programmable nuclease, wherein the composition further comprises a guide RNA associated with the RNA-programmable nuclease,

wherein the supercharged protein has an overall positive charge of at least +25 at pH 7.4, and wherein the supercharged protein has an overall positive charge that is at least +5 greater than its corresponding unmodified protein and is in sufficient quantity and is formulated for penetration into a cell.

2. The composition of claim 1, wherein the RNA-programmable nuclease is a Cas9 nuclease.

3. The composition of any one of claims 1-2, wherein the RNA-programmable nuclease specifically binds and cleaves a nucleic acid sequence.

4. The composition of claim 3, wherein the nucleic acid sequence is comprised in a gene or allele associated with a disease or disorder.

5. The composition of claim 3 or 4, wherein the guide RNA comprises a sequence identical or homologous to the nucleic acid sequence bound and cleaved by the nuclease.

6. The composition of any one of claims 3-5, wherein the nucleic acid sequence is a sequence comprised in a gene that is a therapeutic target.

7. The composition of claim 6, wherein the gene is a gene controlling cell fate.

8. The composition of claim 7, wherein the gene is a gene that induces or inhibits cellular programming towards a brown adipocyte cell fate.

9. The composition of claim 7 or 8, wherein the gene is PRDM16 or PPARγ.

10. The composition of any one of claims 1-9, wherein the supercharged protein is associated with the RNA-programmable nuclease via a covalent bond, thus forming a fusion protein.

11. The composition of any one of claims 1-10, wherein the supercharged protein is associated with the RNA-programmable nuclease via a linker.

12. The composition of claim 11, wherein the linker is a cleavable linker.

13. The composition of claim 12, wherein the linker is a UV-cleavable linker or a linker that is cleaved by a lysosomal enzyme.

14. The composition of any one of claims 1-9, wherein the supercharged protein is associated with the RNA-programmable nuclease via non-covalent interactions, thus forming a complex.

15. The composition of claim 14, wherein the supercharged protein has an overall net positive charge of at least +30 at pH 7.4.

16. The composition of claim 14, wherein the supercharged protein has overall net positive charge of at least +35 at pH 7.4.

17. The composition of claim 14, wherein the overall net positive charge of the complex is about +40.

18. The composition of any one of claims 1-17, wherein the supercharged protein is at least +10 more positively charged at pH 7.4 than its corresponding unmodified protein.

19. The composition of any one of claims 1-17, wherein the supercharged protein is at least +15 more positively charged at pH 7.4 than its corresponding unmodified protein.

20. The composition of any one of claims 1-19, wherein the supercharged protein is a fluorescent protein.

21. The composition of any one of claims 1-20, wherein the supercharged protein is green fluorescent protein (GFP).

22. The composition of any one of claims 1-21, wherein the supercharged protein is a superpositively charged GFP.

23. The composition of any one of claims 1-22, wherein the supercharged protein is a superpositively charged GFP comprising at least 20 contiguous amino acid residues of the sequence:
GGASKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKLPVPW
PTLVTTLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEG
RTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGIKAKFKIRHNVKDGSV
QLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKRDHMVLLEFVTAAGIKHGRDE
RYK (SEQ ID NO: 1).

24. The composition of claim 23, wherein the supercharged protein comprises or consists of the amino acid sequence set forth in SEQ ID NO: 1.

25. The compositions of any one of claims 1-24, wherein the composition is a pharmaceutical composition.

26. The pharmaceutical composition of claim 25, wherein the composition comprises a pharmaceutically acceptable excipient.

27. The pharmaceutical composition of claim 25 or 26, wherein the composition is formulated for administration to a subject and comprises the supercharged protein and the RNA-programmable nuclease in an amount effective for delivery to at least one cell of the subject.

28. The pharmaceutical composition of any one of claims 25-27, wherein the composition comprises the supercharged protein and the RNA-programmable nuclease in an amount effective for inducing a measurable therapeutic effect after administration to a subject.

29. Use of composition of any one of claims 1-24 or pharmaceutical composition of any one of claims 25-28 in the manufacture of a medicament for the treatment of genetic disease, optionally wherein the genetic disease is associated with abnormally elevated levels of an mRNA, a protein, or combination thereof.

30. The use of claim 29, wherein the composition comprises a RNA-programmable nuclease that specifically binds and cleaves a gene encoding a pathogenic RNA or protein, or a gene encoding an RNA or protein that is expressed at abnormally high levels in diseased cells or tissue, optionally wherein the medicament is to be administered under conditions sufficient for the RNA-programmable nuclease to penetrate a cell of a subject, and optionally wherein administration of the medicament comprises a route of administration selected from the group consisting of oral, intravenous, intramuscular, intra-arterial, subcutaneous, intraventricular, topical, inhalational, and mucosal delivery.

31. An ex-vivo/in-vitro method of introducing an RNA-programmable nuclease into a cell, the method comprising contacting the cell with the composition of any one of claims 1-24 or with the pharmaceutical composition of any one of claims 25-28, under conditions suitable for the RNA-programmable nuclease to enter the cell, thereby introducing the RNA-programmable nuclease into the cell, optionally wherein the cell is a somatic cell, and optionally wherein the cell is contacted with the composition or the pharmaceutical composition in an amount, for a time, and under conditions sufficient to induce programming of the cell to a desired cell fate.

32. The method of claim 31, further comprising confirming that the RNA-programmable nuclease has penetrated the cell, optionally wherein the programmed cell is to be used in a cell replacement therapeutic approach, and optionally wherein the cell is contacted *ex vivo* and is to be re-administered to the subject after successful targeting of the undesired allele by the nuclease.

33. The method of claim 31 or 32, wherein the cell is obtained from a subject diagnosed with having or being at risk of developing a disorder associated with an abnormal expression level of a gene, and wherein the RNA-programmable nuclease modulates the expression level of the gene, optionally wherein the cell is a cell carrying a genomic allele associated with a disease, and the RNA-programmable nuclease is a nuclease specifically targeting the allele.

34. A kit comprising the composition of any one of claims 1-24 or the pharmaceutical composition of any one of claims 25-28.

35. A kit for carrying out the method of any one of claims 31-33, wherein the kit comprises the composition of any one of claims 1-24, or the pharmaceutical composition of any one of claims 25-28.

36. The composition of any one of claims 1-24 or the pharmaceutical composition of any one of claims 25-28 for use in medicine.