ENGINEERED CASCADE COMPONENTS AND CASCADE COMPLEXES

Cross-Reference to Related Applications

[0001] This application is a continuation-in-part of U.S. Patent Application Serial No. 16/420,061, filed 22 May 2019, now pending, which is a continuation of U.S. Patent Application Serial No. 16/262,773, filed 30 January 2019, now allowed, which is a continuation of U.S. Patent Application Serial No. 16/104,875, filed 17 August 2018, now U.S. Patent No. 10,227,576, issued 12 March 2019, and claims the benefit of U.S. Provisional Patent Application Serial No. 62/684,735, filed 13 June 2018, now pending, and U.S. Provisional Patent Application Serial No. 62/807,717, filed 19 February 2019, now pending: the contents of which are herein incorporated by reference in their entireties.

Statement Regarding Federally Sponsored Research or Development

[0002] Not applicable.

Sequence Listing

[0003] The present application contains a Sequence Listing that has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. The ASCII copy, created on 12 June 2019 is named CBI032-30_ST25.txt and is 3.1 MB in size.

Technical Field

[0004] The present disclosure relates generally to engineered Class 1 Type I CRISPR-Cas (Cascade) systems that comprise multi-protein effector complexes, nucleoprotein complexes comprising Type I CRISPR-Cas subunit proteins and nucleic acid guides, polynucleotides encoding Type I CRISPR-Cas subunit proteins, and guide polynucleotides. The disclosure also relates to compositions and methods for making and using the engineered Type I CRISPR-Cas systems of the present invention.

Background

[0005] Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) constitute CRISPR-Cas systems. The CRISPR-Cas systems provide adaptive immunity against foreign polynucleotides in bacteria and archaea (see, e.g., Barrangou, R., et al., Science 315:1709-1712 (2007); Makarova, K.S., et al., Nature Reviews Microbiology 9:467-477 (2011); Garneau, J.E., et al., Nature 468:67-71 (2010); Sapranauskas, R., et al., Nucleic Acids Res. 39:9275-9282 (2011); Koonin, E.V., et al., Curr. Opin. Microbiol. 37:67-78 (2017)). Various CRISPR-Cas systems in their native hosts are capable of DNA targeting (Class 1 Type I; Class 2 Type II and Type V), RNA targeting (Class 2 Type VI), and joint DNA and RNA targeting (Class 1 Type III) (see, e.g., Makarova, K.S., et al., Nat. Rev.

Microbiol. 13:722-736 (2015); Shmakov, S., et al., Nat. Rev. Microbiol. 15:169-182 (2017); Abudayyeh, O.O., et al., Science 353:1-17 (2016)).

[0006] The classification of CRISPR-Cas systems has had many iterations. Koonin, E.V., et al., (Curr. Opin. Microbiol. 37:67-78 (2017)) proposed a classification system that takes into consideration the signature *cas* genes specific for individual types and subtypes of CRISPR-Cas systems. The classification also considered sequence similarity between multiple shared Cas proteins, the phylogeny of the best conserved Cas protein, gene organization, and the structure of the CRISPR array. This approach provided a classification scheme that divides CRISPR-Cas systems into two distinct classes: Class 1 comprising a multiprotein effector complex (Type I (CRISPR-associated complex for antiviral defense ("Cascade") effector complex), Type III (Cmr/Csm effector complex), and Type IV); and Class 2 comprising a single effector protein (Type II (Cas9), Type V (Cas12a, previously referred to as Cpf1), and Type VI (Cas13a, previously referred to as C2c2)). In the Class 1 systems, Type I is the most common and diverse, Type III is more common in archaea than bacteria, and Type IV is least common.

[0007] The Type I systems comprise the signature Cas3 protein. The Cas3 protein has helicase and DNase domains responsible for DNA target sequence cleavage. To date, seven subtypes of the Type I system have been identified (i.e., Type I-A, I-B, I-C, I-D, I-E, I-F (and variants for I-F (e.g., I-Fv1, I-Fv2)), and I-U) that have a variable number of *cas* genes. Type I *cas* genes include, but are not limited to, the following: *cas7, cas5, cas8, cse2, csa5, cas3, cas2, cas4, cas1,* and *cas6*. Examples of organisms having Type I systems are as follows: I-A, *Archaeoglobus fulgidus;* I-B, *Clostridium kluyveri;* I-C, *Bacillus halodurans;* I-U, *Geobacter sulfurreducens;* I-D, *Cyanothece sp.* 8802*;* I-E, *Escherichia coli* K12 (*E. coli* K12); I-F, *Yersinia pseudo-tuberculosis;* I-F variant, *Shewanella putrefaciens* CN-32 (Koonin, E.V., et al., Curr. Opin. Microbiol. 37:67-78 (2017)). Characteristics of Cas3 protein mediated cleavage and progressive degradation of DNA have been described (see, e.g., Plagens, A., et al., Nucleic Acids Res. 42:5125-5138 (2014); Maier, L., et al., RNA Biol. 10:865–874 (2013); Hochstrasser, M., et al., Proc. Natl. Acad. Sci. USA 111:6618–6623 (2014); Sinkunas, T., et al., EMBO J. 30:1335–1342 (2011); Westra, E., et al., Mol. Cell 46:595–605 (2012); Mulepati, S., et al., J. Biol. Chem. 288:22184–22192 (2013); Sinkunas, T., et al., EMBO J. 32:385–394 (2013); Mulepati, S., et al., J. Biol. Chem. 288:22184–22192 (2013); Redding, S., et al., Cell 163:854–865 (2015); Sinkunas, T., et al., EMBO J. 32:385–394 (2013); Westra, E., et al., Mol. Cell 46:595–605 (2012)).

[0008] Type I systems typically encode proteins that combine with a CRISPR RNA (crRNA or "guide RNA") to form a Cascade complex. These complexes comprise multiple proteins and a crRNA, both of which are transcribed from this CRISPR locus. In Type I systems, primary processing of a pre-crRNA is catalyzed by Cas6. This typically results in a crRNA with a 5′ handle of 8 nucleotides, a spacer region, and a 3′ handle; both the 5' and the 3' handles are derived from the repeat sequence. In some systems, the 3′ handle forms a stem-loop structure; in other systems, secondary processing of the 3′ end of crRNA is catalyzed by ribonuclease(s) (see, e.g., van der Oost, J., et al., Nature Reviews Microbiology 12:479-492 (2014)).

[0009] The Cascade effector complexes of the Type I CRISPR-Cas systems comprise a backbone having paralogous Repeat-Associated Mysterious Proteins (RAMPs; e.g., Cas7 and Cas5 proteins) containing the RNA Recognition Motif (RRM) fold and additional "large" and "small" subunit proteins (see, e.g., Koonin, E.V., et al., Curr. Opin. Microbiol. 37:67-78, (2017), Figure 2). These Cascade effector complexes typically have a Cas5 subunit protein and several Cas7 subunit proteins. Such Cascade effector complexes also comprise the guide RNA. The Cascade effector complexes comprise the various subunit proteins arranged in an asymmetric fashion along the length of the guide RNA. The Cas5 subunit protein and the large subunit protein (Cas8 protein) are positioned at one end of the complex, enveloping the 5' end of the guide RNA. Several copies of the small subunit protein interact with the guide RNA backbone, which is bound to multiple copies of the Cas7 subunit protein. The Cas6 subunit protein, another RAMP protein, is associated with the Cascade effector complex primarily through association with the 3′ handle (repeat region) of the crRNA. The Cas6 subunit protein usually functions as the repeat-specific RNase involved in pre-crRNA processing; however, in Type I-C systems, Cas5 functions as the repeat-specific RNase and there is no Cas6.

[0010] The primary sequences of the CRISPR-Cas Type I Cascade subunit proteins have little sequence identity; however, the presence of homologous RAMP modules and the overall structural similarity of the multiprotein effector complexes supports a common origin of these effector complexes (see, e.g., Koonin, E.V., et al., Curr. Opin. Microbiol. 37:67-78 (2017)).

[0011] The adaptive immunity mechanism of action in the Type I CRISPR-Cas systems involves essentially three phases: adaptation, expression, and interference. In the adaptation phase, a foreign DNA or RNA infects the host and proteins encoded by various *cas* genes bind regions of the infecting DNA or RNA. Such regions are called protospacers. A protospacer adjacent motif (PAM) is a short nucleotide sequence (e.g., 2 to 6 base pair DNA sequence) that

is adjacent to the protospacer. PAM sequences are typically recognized by a Cas1 subunit protein/Cas2 subunit protein complex, wherein the active PAM-sensing site is associated with the Cas1 subunit proteins (see, e.g., Jackson, S. A., et al., Science 356:356(6333) (2017)).

[0012] In the expression phase, the CRISPR array comprising multiple spacer-repeat elements is transcribed as a single transcript. Individual spacer repeat elements are processed by an endonuclease (e.g., Type I, a Cas6 protein; and Type I-C, a Cas5 protein) into individual crRNAs. Cas subunit proteins are expressed and associate with the crRNA to form a Cascade effector complex.

[0013] The Cascade effector complex scans foreign polynucleotides infecting the host to identify DNA complementary to the spacer. In Type I systems, interference occurs when the effector complex identifies a sequence complementary to the spacer that is adjacent to a PAM; and the Cas3 protein is recruited to the DNA-bound Cascade effector complex to cleave and progressively digest the foreign polynucleotide.

[0014] Makarova, K.S., et al., (Cell 168:946 (2017)) provide a summary of genes, homologs, Cascade complexes, and mechanisms of action for Type I CRISPR-Cas systems.

[0015] Type I CRISPR-Cas systems have thus far had limited use in eukaryotic genome engineering applications, due in part to the difficulty of heterologous expression of the Cascade complex and the way in which the Type I CRISPR-Cas systems cleave DNA targets.

Summary of the Invention

[0016] The present invention generally relates to compositions comprising engineered Type I CRISPR-Cas effector complexes and components thereof, including protein components, modified or distinctly changed guide polynucleotides, and combinations thereof.

[0017] One embodiment of the present invention is a composition comprising:

[0018] a first engineered Type I CRISPR-Cas effector complex comprising,

[0019] a first Cse2 subunit protein, a first Cas5 subunit protein, a first Cas6 subunit protein, and a first Cas7 subunit protein,

[0020] a first fusion protein comprising a first Cas8 subunit protein and a first FokI, wherein the N-terminus of the first Cas8 subunit protein or the C-terminus of the first Cas8 subunit protein is covalently connected by a first linker polypeptide to the C-terminus or N-terminus, respectively, of the first FokI, and wherein the first linker polypeptide has a length of between 10 amino acids and 40 amino acids, and

[0021] a first guide polynucleotide comprising a first spacer capable of binding a first nucleic acid target sequence; and

[0022] a second engineered Type I CRISPR-Cas effector complex comprising,

[0023] a second Cse2 subunit protein, a second Cas5 subunit protein, a second Cas6 subunit protein, and a second Cas7 subunit protein,

[0024] a second fusion protein comprising a second Cas8 subunit protein and a second FokI, wherein the N-terminus of the second Cas8 subunit protein or the C-terminus of the second Cas8 protein is covalently connected by a second linker polypeptide to the C-terminus or N-terminus, respectively, of the second FokI, and wherein the second linker polypeptide has a length of between 10 amino acids and 40 amino acids, and

[0025] a second guide polynucleotide comprising a second spacer capable of binding a second nucleic acid target sequence, wherein a protospacer adjacent motif (PAM) of the second nucleic acid target sequence and a PAM of the first nucleic acid target sequence have an interspacer distance between 20 base pairs and 42 base pairs.

[0026] In some embodiments, the length of the first linker polypeptide and/or the second linker polypeptide is a length of between 15 amino acids and 30 amino acids, or between 17 amino acids and 20 amino acids. In one embodiment, the length of the first linker polypeptide and the second linker polypeptide are the same.

[0027] Interspacer distances between the second nucleic acid target sequence and the first nucleic acid target sequence include, but are not limited to, between 22 base pairs and 40 base pairs, between 26 base pairs and 36 base pairs, between 29 base pairs and 35 base pairs, or between 30 base pairs and 34 base pairs.

[0028] The first FokI and the second FokI can be monomeric subunits that are capable of associating to form a homodimer, or distinct subunits that are capable of associating to form a heterodimer.

[0029] In some embodiments, the N-terminus of the first Cas8 subunit protein is covalently connected by the first linker polypeptide to the C-terminus of the first FokI, the C-terminus of the first Cas8 subunit protein is covalently connected by a first linker polypeptide to the N-terminus of the first FokI, the N-terminus of the second Cas8 subunit protein is covalently connected by the second linker polypeptide to the C-terminus of the second FokI, the C-terminus of the second Cas8 subunit protein is covalently connected by a second linker polypeptide to the N-terminus of the second FokI, and combinations thereof. The first Cas8 subunit protein and the second Cas8 subunit protein can each comprise a Cas8 subunit protein having a different sequence or both the first and the second Cas8 subunit protein can comprise identical amino acid sequences.

[0030] Similarly, the first Cse2 subunit protein and the second Cse2 subunit protein can each comprise different or identical Cse2 subunit protein amino acid sequences, the first Cas5

subunit protein and the second Cas5 subunit protein can each comprise different or identical Cas5 subunit protein amino acid sequences, the first Cas6 subunit protein and the second Cas6 subunit protein can each comprise different or identical Cas6 subunit protein amino acid sequences, the first Cas7 subunit protein and the second Cas7 subunit protein can each comprise different or identical Cas7 subunit protein amino acid sequences, and combinations thereof.

[0031] In a preferred embodiment, the guide polynucleotides comprise RNA.

[0032] In an additional embodiment, the present invention includes an engineered Type I CRISPR Cas3 mutant protein ("mCas3 protein") capable of reduced movement along DNA relative to a wild-type Type I CRISPR Cas3 protein ("wtCas3 protein").

[0033] The present invention also includes the use of the above compositions to perform genome editing in cells, as well as methods of make the above compositions.

[0034] Further embodiments of the present invention will be readily apparent to those of ordinary skill in the art in view of the disclosures herein.

Brief Description of the Figures

[0035] The Figures are not proportionally rendered, nor are they to scale. The locations of indicators are approximate.

[0036] **FIG. 1A** present a generalized illustration of a Type I CRISPR-Cas effector complex. **FIG. 1B** presents a generalized illustration of a Type I CRISPR-Cas crRNA.

[0037] **FIG. 2A**, **FIG. 2B**, and **FIG. 2C** present illustrative examples of two engineered Type I CRISPR-Cas effector complexes with fusion domains bound to neighboring spacer sequences.

[0038] **FIG. 3A** and **FIG. 3B** present examples of circularly permuted proteins.

[0039] **FIG. 4A**, **FIG. 4B**, **FIG. 5A**, **FIG. 5B**, **FIG. 6A**, **FIG. 6B**, **FIG. 6C**, **FIG. 7A**, **FIG. 7B**, **FIG. 8**, **FIG. 9**, **FIG. 10A**, and **FIG. 10B** illustrate a variety of examples of engineered Type I CRISPR-Cas effector complexes of the present invention.

[0040] **FIG. 11A** and **FIG. 11B** illustrate examples of substrate channels.

[0041] **FIG. 12A**, **FIG. 12B**, and **FIG. 12C** present a generalized illustration of site-directed recruitment of a functional protein domain fused to a Cascade subunit protein by a dCas9:NATNA complex.

[0042] **FIG. 13A**, **FIG. 13B**, **FIG. 14A**, **FIG. 14B**, and **FIG. 14C** illustrate examples of engineered Type I CRISPR-Cas effector complexes of the present invention.

[0043] **FIG. 15A**, **FIG. 15B**, **FIG. 15C**, **FIG. 16A**, **FIG. 16B**, **FIG. 16C**, **FIG. 17A**, **FIG. 17B**, **FIG. 17C**, **FIG. 18A**, **FIG. 18B**, **FIG. 18C**, **FIG. 18D**, **FIG. 19A**, **FIG. 19B**, **FIG.**

20A, and **FIG. 20B** present examples of engineered Type I CRISPR-Cas effector complexes of the present invention and methods of use thereof.

[0044] **FIG. 21A**, **FIG. 21B**, **FIG. 21C**, **FIG. 21D**, **FIG. 22A**, **FIG. 22B**, **FIG. 22C**, and **FIG. 22D** illustrate embodiments of the present invention that use a Cas3 protein comprising active endonuclease activity.

[0045] **FIG. 23A**, **FIG. 23B**, **FIG. 23C**, **FIG. 23D**, **FIG. 23E**, **FIG. 24**, **FIG. 25**, **FIG. 26**, and **FIG. 27** present schematic diagrams of a variety of Cascade component expression systems.

[0046] **FIG. 28**, **FIG. 29**, **FIG. 30**, **FIG. 31A**, **FIG. 31B**, **FIG. 32**, **FIG. 33A**, **FIG. 33B**, and **FIG. 34** present data related to genome editing of the engineered Cascade systems of the present invention.

[0047] **FIG. 35** illustrates an example of a minimal CRISPR array containing paired guide RNAs (gRNAs).

[0048] **FIG. 36A, FIG. 36B, FIG. 36C,** and **FIG. 36D** present data related to genome editing in human cells via RNP and plasmid-based delivery of engineered Type I CRISPR-Cas complexes.

[0049] **FIG. 37A, FIG. 37B, FIG. 37C, FIG. 37D, FIG. 37E**, **FIG. 37F**, and **FIG. 37G** present data related to repair outcomes.

[0050] **FIG. 38A, FIG. 38B,** and **FIG. 38C** present data related to how mismatches between gRNAs and target DNA inhibit genome editing by engineered Type I CRISPR-Cas complexes.

[0051] **FIG. 39A, FIG. 39B, FIG. 39C,** and **FIG. 39D** presents data related to expanded screening of PAM selectivity for three Cascade homolog variants.

[0052] **FIG. 40A, FIG. 40B, FIG. 40C, FIG. 40D, FIG. 40E**, and **FIG. 40F** present data related to exemplary changes in editing efficiency of engineered Type I CRISPR-Cas complexes.

[0053] **FIG. 41A, FIG. 41B,** and **FIG. 41C** present data related to expanded screening of FokI-Cas8 linker length and interspacer distance for three Cascade homolog variants.

[0054] **FIG. 42A** and **FIG. 42B** illustrate an example of oligo-templated PCR amplification.

[0055] **FIG. 43** presents data for percent genome editing is shown as a function of FokI-Cascade homolog variant and interspacer distance.

[0056] **FIG. 44** shows a linear representation of the functional domains of the EcoCas3 protein and the relative locations of mutants made within the sequence.

[0057] **FIG. 45A**, **FIG. 45B**, **FIG. 45C**, and **FIG. 45D** show data related to genome editing using EcoCascade RNP complexes comprising wild-type or mutant EcoCas3 proteins.

[0058] **FIG. 46A**, **FIG. 46B**, **FIG. 46C**, **FIG. 47A**, and **FIG. 47B** present data related to dCas9-VP64/sgRNA RNP complex roadblocks and their effect on cleavage of targets by EcoCascade RNP complexes.

[0059] **FIG. 48** show exemplary editing data for Cas3[D452A]/-EcoCascade or mCas3[D452A]–EcoCascade.

[0060] **FIG. 49** presents data for genome editing at eight TRAC target sites with PseCascade RNP complexes.

Incorporation by Reference

[0061] All patents, publications, and patent applications cited in the present Specification are herein incorporated by reference as if each individual patent, publication, or patent application was specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

Detailed Description of the Invention

[0062] It is to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in the present Specification and the Claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes one or more polynucleotides, and reference to "a vector" includes one or more vectors.

[0063] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although other methods and materials similar, or equivalent, to those described herein can be useful in the present invention, preferred materials and methods are described herein.

[0064] In view of the teachings of the present Specification and the Examples, one of ordinary skill in the art can apply conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics, and recombinant polynucleotides, as taught, for example, by the following standard texts: Cellular and Molecular Immunology, Ninth Edition, A. K. Abbas., et al., Elsevier (2017), ISBN 978-0323479783; Cancer Immunotherapy Principles and Practice, First Edition, L.H. Butterfield, et al., Demos Medical (2017), ISBN 978-1620700976; Janeway's Immunobiology, Ninth Edition, Kenneth

Murphy, Garland Science (2016), ISBN 978-0815345053; Clinical Immunology and Serology: A Laboratory Perspective, Fourth Edition, C. Dorresteyn Stevens, et al., F.A. Davis Company (2016), ISBN 978-0803644663; Antibodies: A Laboratory Manual, Second edition, E.A. Greenfield, Cold Spring Harbor Laboratory Press (2014), ISBN 978-1-936113-81-1; Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, Seventh Edition, R.I. Freshney, Wiley-Blackwell (2016), ISBN 978-1118873656; Transgenic Animal Technology, Third Edition: A Laboratory Handbook, C.A. Pinkert, Elsevier (2014), ISBN 978-0124104907; The Laboratory Mouse, Second Edition, H. Hedrich, Academic Press (2012), ISBN 978-0123820082; Manipulating the Mouse Embryo: A Laboratory Manual, Fourth Edition, R. Behringer, et al., Cold Spring Harbor Laboratory Press (2013), ISBN 978-1936113019; PCR 2: A Practical Approach, M.J. McPherson, et al., IRL Press (1995), ISBN 978-0199634248; Methods in Molecular Biology (Series), J.M. Walker, ISSN 1064-3745, Humana Press; RNA: A Laboratory Manual, D.C. Rio, et al., Cold Spring Harbor Laboratory Press (2010), ISBN 978-0879698911; Methods in Enzymology (Series), Academic Press; Molecular Cloning: A Laboratory Manual (Fourth Edition), M.R. Green, et al., Cold Spring Harbor Laboratory Press (2012), ISBN 978-1605500560; Bioconjugate Techniques, Third Edition, G.T. Hermanson, Academic Press (2013), ISBN 978-0123822390; Methods in Plant Biochemistry and Molecular Biology, W.V. Dashek, CRC Press (1997), ISBN 978-0849394805; Plant Cell Culture Protocols (Methods in Molecular Biology), V.M. Loyola-Vargas, et al., Humana Press (2012), ISBN 978-1617798177; Plant Transformation Technologies, C.N. Stewart, et al., Wiley-Blackwell (2011), ISBN 978-0813821955; Recombinant Proteins from Plants (Methods in Biotechnology), C. Cunningham, et al., Humana Press (2010), ISBN 978-1617370212; Plant Genomics: Methods and Protocols (Methods in Molecular Biology), W. Busch, Humana Press (2017), ISBN 978-1493970018; Plant Biotechnology: Methods in Tissue Culture and Gene Transfer, R. Keshavachandran, et al., Orient Blackswan (2008), ISBN 978-8173716164.

[0065] Clustered regularly interspaced short palindromic repeats (CRISPR) and related CRISPR-associated proteins (Cas proteins) constitute CRISPR-Cas systems (see, e.g., Barrangou, R., et al., Science 315:1709-1712 (2007)).

[0066] As used herein, "Cas protein," "CRISPR-Cas protein," and "CRISPR-Cas subunit protein," and "Cas subunit protein," unless otherwise identified, all refer to Class 1 Type I CRISPR-Cas proteins. Typically, for use in aspects of the present invention, Cas subunit proteins are capable of interacting with one or more cognate polynucleotides (most typically, a crRNA) to form a Type I effector complex (most typically, an RNP complex).

[0067] The genes encoding Cascade in Type I-E CRISPR–Cas systems have been named with various conventions over time, which may serve as a point of confusion when comparing recent and older literature. Typically, the present Specification uses the nomenclature as set forth in Koonin, E., et al. (Curr. Opin. Microbiol. 37: 67–78 (2017)), in which the gene order in the reference *E. coli* K12 operon is: *cas3, cas8, cas11, cas7, cas5, cas6, cas1,* and *cas2*. For simplicity's sake, the "*e*" qualifier in *cas8e* is sometimes used to distinguish the *cas8* gene between different subtypes within Type I systems. The stoichiometry of the wild-type *E. coli* Type I-E CRISPR–Cas is $Cas5_1$-$Cas6_1$-$Cas7_6$-$Cas8_1$-$Cas11_2$-$gRNA_1$.

[0068] However, for the purposes of cross-referencing: *cas8* has been previously referred to as *cse1* and *casA*, and also known as the "large subunit"; *cas11* has been previously referred to as *cse2* and *casB*, and also known as the "small subunit"; *cas7* has been previously referred to as *cse4* and *casC;*, *cas5* has been previously referred to as *casD*, and sometimes given the qualifier *cas5e;* and *cas6* has been previously referred to as *cse3* and *casE*, and often given the qualifier *cas6e*. Genes encoding Cas subunit proteins are listed in Table 1.

[0069]

Table 1
Type I CRISPR-Cas Proteins

Function	Universal family name*	Alternative designation	Reported stoichiometry (when present)
RNA 5' cap, PAM recognition, duplex unwinding	Cas5	CasD, Cas5e, Csc1, Csy2, Csf3, Cas1822	1
PAM recognition, duplex unwinding, Cas3 recruitment	Cas8	Large subunit, CasA, Cse1, Cas8a, Cas8b, Cas8c, Cas8e, Cas8f, Csy1	1
R-loop stabilization	Cas11	Small subunit, CasB, Cse2	2
Backbone	Cas7	CasC, Cse4, Csc2, Csy3, Csf2, Cas1821, Cst2/DevR	3-6
RNA 3' cap	Cas6	CasE, Cse3, Cas6e, Cas6f, Csy4	1
DNA cleavage	Cas3	Cas3', Cas3''	1

*As defined by Makarova, K.S., et al., Nat. Rev. Microbiol. 13:722-736 (2015); Koonin, E.V., et al., Curr Opin Microbiol. 37:67-78 (2017).

[0070] PAM sequences are typically recognized by a Cas1 subunit protein/Cas2 subunit protein complex, wherein the active PAM-sensing site is associated with the Cas1 subunit proteins (see, e.g., Jackson, S. A., et al., Science 356:356(6333) (2017)). Cas1 protein and Cas2 protein are present in the great majority of the known CRISPR-Cas systems and are sufficient for the insertion of spacers into CRISPR cassettes (see, e.g., Yosef, I, et al., Nucleic Acids Res.

40:5569–5576 (2012)). These two proteins form a complex for the adaptation process. The endonuclease activity of Cas1 protein is required for spacer integration whereas Cas2 protein appears to perform a non-enzymatic function (see, e.g., Nunez, J., et al., Nat Struct Mol Biol. 21:528–534 (2014); Richter, C., et al., PLoS One. 2012;7:e49549). The Cas1-Cas2 protein complex represents a highly conserved information processing module of CRISPR-Cas systems that appears to be quasi-autonomous from the rest of the system (see, e.g., Makarova, K., et al., Methods Mol. Biol. 1311:47-75 (2015)). The endonuclease Cas1 protein is an essential Cas protein that ensures the unique ability of CRISPR systems to keep memory of previous encounters with infectious agents.

[0071] The terms "Type I CRISPR-Cas effector complex," "Type I CRISPR-Cas nucleoprotein (NP) complex," "Cascade nucleoprotein (NP) complex," and "Type I nucleoprotein (NP) complex," are used interchangeably herein and typically refer to Cascade protein forming a complex with a guide polynucleotide. "Cascade complex" and "Type I complex," are typically used when referring to the protein components of a Cascade NP complex. The terms "Cascade RNP complex," "Type I CRISPR-Cas RNP complex," and "Type I RNP complex," refer to a Cascade complex comprising a crRNA versus a more generic guide polynucleotide (i.e., as in a Cascade NP complex). An example of a wild-type Type I CRISPR-Cas effector complex is illustrated in FIG. 1A. FIG. 1A is adapted from Makarova, K.S., et al., (Cell 168:946 (2017); Makarova, K., et al., Nature Reviews Microbiology 13:722-736 (2015)). FIG. 1A illustrates six Cas7 proteins, a Cas5 protein, a Cas8 protein, two Cse2 proteins, a Cas6 protein, and a crRNA (FIG. 1A: Cas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin) associated as a Cascade complex. The complex is capable of binding a nucleic acid target sequence. After association of a wtCas3 protein (FIG. 1A, Cas3 surrounded by a dashed box) with the complex, the Cascade complex is capable of cleavage of a nucleic acid target sequence. As noted in Table 1, the total number of some Cas subunit proteins can vary in Cascade complexes.

[0072] "Cas3" and "Cas3 protein" are used interchangeably herein to refer to Type I CRISPR-Cas3 proteins, modifications, and variants thereof. The Type I CRISPR-Cas effector complexes bind foreign DNA complementary to the crRNA guide and recruit Cas3, a trans-acting nuclease-helicase required for target degradation. Cas3 proteins have motifs characteristic of helicases from superfamily 2 and contain a DEAD/DEAH box region and a conserved C-terminal domain. Cas3 proteins and variants thereof are known in the art (see, e.g., Westra, E.R., et al., Mol. Cell. 46: 595–605 (2012); Sinkunas, T., et al., EMBO J. 30:1335-

1342 (2011); Beloglazova, N., et al., EMBO J. 30:4616-4627 (2011); Mulepati, S., et al., J. Biol. Chem. 286:31896-31903 (2011)). As used herein, the term "mCas3 protein" refers to a Cas3 protein comprising one or more mutations relative to its corresponding wtCas3 protein. mCas3 proteins include, but are not limited to, mCas3 proteins (e.g., Example 23A, Example 23B, and Example 23C), dblmCas3 proteins (e.g., Example 26A, Example 26B, and Example 26C), and dCas3* (a mutated Cas3 protein that does not have any nuclease activity and/or helicase activity).

[0073] The term "nuclease," as used herein, refers to an enzyme capable of cleaving the phosphodiester bonds, such as those connecting two nucleotides, as found in double-stranded (ds) nucleic acids (e.g., dsDNA, genomic DNA (gDNA), dsRNA), single-stranded (ss) nucleic acids (e.g., ssDNA, RNA) or hybrid dsRNA/DNA. An "endonuclease" typically can affect ss-(nicks) or ds-breaks in its target molecules. One example of a DNA endonuclease is a FokI enzyme. "FokI endonuclease" and "FokI" are used interchangeably herein and refer to a FokI enzyme, FokI homologs, enzymatically active domain(s) of FokI enzymes, and variants of FokI enzymes. FokI dimerization is typically required for DNA cleavage. Dimers of FokI can comprise two monomeric subunits that associate to form a homodimer or two distinct monomeric subunits that associate to form a heterodimer (see, e.g., Bitinaite, J., et al., Proc. Natl. Acad. Sci. USA 95:10570-10575 (1998); Ramalingam, S., et al., J. Mol. Biol. 405:630-641 (2011)). One example of a FokI variant is the *Sharkey* variant described by Guo, et al. (Guo, J., et al., J. Mol. Biol. 400:96-107 (2010)). Additional DNA and RNA nucleases are known in the art.

[0074] "CRISPR RNA," "crRNA," and "guide RNA," as used herein, refer to one or more RNAs with which Cas subunit proteins are capable of interacting to form a Type I effector complex that guides the complex to preferentially bind a nucleic acid target sequence in a polynucleotide (relative to a polynucleotide that does not comprise the nucleic acid target sequence). "Guide" and "guide polynucleotide," as used herein, refer to a polynucleotide component of Type I effector complexes comprising ribonucleotide bases (e.g., RNA) and ribose sugars, as well as disparate components and combinations thereof, including but not limited to deoxyribonucleotide bases, nucleotide analogs, modified nucleotides, different nitrogenous bases, fundamentally different nucleotide bases, chemically disparate molecules, intermixtures of bases (e.g., RNA bases, DNA bases, and/or modified bases), and the like as well as combinations thereof, in addition to synthetic backbones, naturally occurring backbones, non-naturally occurring backbones, fundamentally different backbone residues, chemically disparate residues or linkages, modified backbones, intermixtures (e.g., ribose and

deoxyribose components of a backbone), and the like, as well as combinations thereof. Some examples of guide polynucleotides are described herein. An example of a Type I CRISPR-Cas crRNA associated with a nucleic acid target sequence through the crRNA spacer is illustrated in FIG. 1B. FIG. 1B is adapted from Hochstrasser, M.L., et al., Mol. Cell 63:840-851 (2016). In FIG. 1B, the PAM (FIG. 1B, 104) is associated with the nucleic acid target sequence and the 5' and 3' strands of a double-stranded nucleic acid are illustrated (FIG. 1B, vertical lines represent hydrogen bonds). A guide polynucleotide (FIG. 1B, 106) typically comprises a 5' handle region (FIG. 1B, 101), a spacer region (FIG. 1B, 103) comprising a seed region, and a 3' hairpin comprising two hydrogen-bonded repeat regions (FIG. 1B, 102); horizontal lines represent hydrogen bonds. PAM sequences associated with a number of Type I Cascade homologs are discussed herein. The PAM sequences are adjacent protospacer sequences (FIG. 1B, 105). FIG. 1B illustrates the Cascade complex spacer bound to the nucleic acid target sequence (FIG. 1B, vertical lines represent hydrogen bonds). FIG. 1B also illustrates the protospacer region (FIG. 1B, protospacer). The spacer can comprise a region of the crRNA between about 6 and about 56 nucleotides, wherein the spacer is complementary to a nucleic acid target sequence in a polynucleotide. The spacer length can be changed to fine-tune Cascade activity in Type I-E CRISPR-Cas systems. Cascade complexes can incorporate an extra Cas7 subunit with every 6 nucleotides added to the crRNA spacer and an extra Cse2 subunit with every 12 nucleotides added to the spacer (see, e.g., Luo, M.L., et al., Nucleic Acids Res. 44(15):7385-7394 (2016)). The spacer typically comprises a region of between about 32 and about 36 nucleotides.

[0075] The terms "spacer," "spacer sequence," and "nucleic acid target binding sequence" are used interchangeably herein.

[0076] "Target," "target sequence," "nucleic acid target sequence," and "on-target sequence" are used interchangeably herein to refer to a nucleic acid sequence that is wholly, or in part, complementary to a nucleic acid target binding sequence of the guide (e.g., the spacer of a crRNA) of a Cascade nucleoprotein complex (e.g., a Cascade RNP complex). Typically, the nucleic acid target binding sequence is selected to be 100% complementary to a nucleic acid target sequence to which binding of a Cascade nucleoprotein complex is being directed; however, to attenuate binding to a nucleic acid target sequence, lower percent complementarity can be used. When the target binding sequence is 100% complementary to the target sequence, "off-target" sequence binding refers to binding of the Cascade nucleoprotein complex to nucleic acid sequences having less than 100% complementarity to the nucleic acid target binding sequence (spacer). A double-stranded DNA sequence typically comprises a nucleic

acid target sequence on one strand (FIG. 1B, section hydrogen bonded to the guide RNA). A "target region" comprises a nucleic acid target sequence.

[0077] As used herein, a "stem element" or "stem structure" refers to two strands of nucleic acids that are known to, or predicted to, form a double-stranded region (the "stem element"). A "stem-loop element" or "stem-loop structure" refers to a stem structure wherein 3'-end sequences of one strand are covalently bonded to 5'-end sequences of the second strand by a nucleotide sequence of typically single-stranded nucleotides ("a stem-loop element nucleotide sequence"). In some embodiments, the loop element comprises a loop element nucleotide sequence of between about 3 and about 20 nucleotides in length, preferably between about 4 and about 10 nucleotides in length. In preferred embodiments, a loop element nucleotide sequence is a single-stranded nucleotide sequence of unpaired nucleic acid bases that do not interact through hydrogen bond formation to create a stem element within the loop element nucleotide sequence. The term "hairpin element" is also used herein to refer to stem-loop structures. Such structures are well known in the art. The base pairing may be exact; however, as is known in the art, a stem element does not require exact base pairing. Thus, the stem element may include one or more base mismatches or non-paired bases. An example of a stem-loop structure in a guide polynucleotide is illustrated in FIG. 1B.

[0078] "Linker element nucleotide sequence," "linker nucleotide sequence," and "linker polynucleotide" are used interchangeably herein and refer to either a single-stranded nucleic acid sequence or a double-stranded nucleic acid sequence of one or more nucleotides covalently attached to a first nucleic acid sequence (e.g., 5'-linker nucleotide sequence-first nucleic acid sequence-3'). In some embodiments, a linker nucleotide sequence connects two separate nucleic acid sequences to form a single polynucleotide (e.g., 5'-first nucleic acid sequence-linker nucleotide sequence-second nucleic acid sequence-3'). Other examples of linker nucleotide sequences include, but are not limited to, 5'-first nucleic acid sequence-linker nucleotide sequence-3' and 5'-linker nucleotide sequence-first first nucleic acid sequence-linker nucleotide sequence-3'. In some embodiments, the linker element nucleotide sequence can be a single-stranded nucleotide sequence of unpaired nucleic acid bases that do not interact with each other through hydrogen bond formation to create a secondary structure (e.g., a stem-loop structure) within the linker element nucleotide sequence. In some embodiments, two linker element nucleotide sequences can interact with each other through hydrogen bonding between the two linker element nucleotide sequences. In some embodiments, a linker polynucleotide encodes a "linker polypeptide." Such a linker polynucleotide typically connects the 3' end of a first polynucleotide encoding a first polypeptide to the 5' end of a second polynucleotide

encoding a second polypeptide to form a single polynucleotide that encodes a fusion protein comprising N-the first polypeptide-the linker polypeptide-the second polypeptide-C. In some embodiments of the present invention, more than two polypeptide sequences can be connected in tandem by linker polypeptides (e.g., N-a first polypeptide-a first linker polypeptide-a second polypeptide-a second linker polypeptide-a third polypeptide-C). "Linker polypeptide," "linker polypeptide sequence," "amino acid linker sequence," and "linker sequence" are also used interchangeably herein.

[0079] As used herein, a "connecting nucleotide sequence" refers to a single-stranded nucleic acid sequence linker sequence that covalently connects a first nucleic acid sequence and a second nucleic acid sequence.

[0080] As used herein, the terms "interspacer," "interspacer region," and "interspacer distance" are interchangeable and refer to the distance between a PAM of a first nucleic acid target sequence (e.g., a first DNA target sequence) and a PAM of a second nucleic acid target sequence (e.g., a second DNA target sequence) typically in a PAM-in orientation, wherein a first Type I CRISPR-Cas effector complex comprises a first spacer capable of binding the first nucleic acid target sequence, and a second Type I CRISPR-Cas effector complex comprises a second spacer capable of binding the second nucleic acid target sequence. FIG. 2A, FIG. 2B, and FIG. 2C present illustrative examples of two Type I CRISPR-Cas effector complexes (FIG. 2A: "Cascade1," solid outlined box, comprising "crRNA1"; and "Cascade2," dashed box, comprising "crRNA2") comprising fusion proteins (FIG. 2A, "FP1" and "FP2" represented as circular sectors; e.g., FP1 and FP can be FokI) connected with each Cascade complex through linker polynucleotides (FIG. 2A, "Linker1" and "Linker2"), wherein the CRISPR-Cas effector complexes are bound to neighboring nucleic acid target sequences on double-stranded DNA (FIG. 2A, "dsDNA," represented as paired, horizontal dashed lines). PAM sequences associated with each nucleic acid target sequence are indicated (FIG. 2A, "PAM1," open box, and "PAM2," open box)). FIG. 2A illustrates an interspacer (shown as a horizontal, double-arrowheaded line at the top of FIG. 2A) between two target sites in a PAM-in (PAM-in/PAM-in) configuration. FIG. 2B illustrates an interspacer (shown as a horizontal, double-arrowheaded line at the top of FIG. 2B) between two target sites in a PAM-in/PAM-out configuration. FIG. 2C illustrates an interspacer (shown as a horizontal, double-arrowheaded line at the top of FIG. 2C) between two target sites in the PAM-out (PAM-out/PAM-out) configuration. FIG. 2A, FIG. 2B, and FIG. 2C also illustrate the separation of the two strands of the dsDNA. A Cascade complex recognizes a dsDNA target sequence adjacent a PAM. PAM sequences are recognized by Cse1. Base pairing between the crRNA and complementary target

DNA strand results in an R-loop with the displaced non-complementary target DNA strand (see, e.g., Beloglazova, N., et al., Nucleic Acids Res. 43:530–543 (2015)).

[0081] As used herein, the term "cognate" refers to biomolecules that interact, such as a cell surface receptor (e.g., a chemokine receptor), and its ligand (e.g., a chemokine expressed on a tumor cell or in a tumor microenvironment); a site-directed polypeptide and its guide; a site-directed polypeptide/guide complex (i.e., a nucleoprotein complex) capable of site-directed binding to a nucleic acid target sequence complementary to the guide binding sequence; and the like. In addition, the term "cognate" refers to a group of Cas subunit proteins (e.g., Cse2, Cas5, Cas6, Cas7, and Cas8) and one or more guide polynucleotides (e.g., a Type I CRISPR-Cas RNA) that are capable of forming a nucleoprotein complex capable of site-directed binding to a nucleic acid target sequence complementary to a spacer present in one of the one or more guide polynucleotides.

[0082] The terms "wild-type," "naturally occurring," and "unmodified" are used herein to mean the typical (or most common) form, appearance, phenotype, or strain existing in nature; for example, the typical form of cells, organisms, polynucleotides, proteins, macromolecular complexes, genes, RNAs, DNAs, or genomes as they occur in, and can be isolated from, a source in nature. The wild-type form, appearance, phenotype, or strain serve as the original parent before an intentional modification, change, mutation, and/or markedly different structural change. Thus, mutant, variant, engineered, recombinant, and modified forms are not wild-type forms.

[0083] The terms "engineered," "genetically engineered," "genetically modified," "recombinant," "modified," "non-naturally occurring," and "non-native" indicate intentional human or machine manipulation of the genome of an organism or cell. The terms encompass methods of genomic modification that include genomic editing, as defined herein, as well as techniques that alter gene expression or inactivation, enzyme engineering, directed evolution, knowledge-based design, random mutagenesis methods, gene shuffling, codon optimization, and the like. Methods for genetic engineering are known in the art.

[0084] "Covalent bond," "covalently attached," "covalently bound," "covalently linked," "covalently connected," and "molecular bond" are used interchangeably herein and refer to a chemical bond that involves the sharing of electron pairs between atoms. Examples of covalent bonds include, but are not limited to, phosphodiester bonds, phosphorothioate bonds, disulfide bonds and peptide bonds (-CO-NH-).

[0085] "Non-covalent bond," "non-covalently attached," "non-covalently bound," "non-covalently linked," "non-covalent interaction," and "non-covalently connected" are used

interchangeably herein and refer to any relatively weak chemical bond that does not involve sharing of a pair of electrons. Multiple non-covalent bonds often stabilize the conformation of macromolecules and mediate specific interactions between molecules. Examples of non-covalent bonds include, but are not limited to, hydrogen bonding, ionic interactions (e.g., Na^+Cl^-), van der Waals interactions, and hydrophobic bonds.

[0086] As used herein, "hydrogen bonding," "hydrogen-base pairing," and "hydrogen bonded" are interchangeable and refer to canonical hydrogen bonding and non-canonical hydrogen bonding including, but not limited to, "Watson-Crick-hydrogen-bonded base pairs" (W-C-hydrogen-bonded base pairs or W-C hydrogen bonding); "Hoogsteen-hydrogen-bonded base pairs" (Hoogsteen hydrogen bonding); and "wobble-hydrogen-bonded base pairs" (wobble hydrogen bonding). W-C hydrogen bonding, including reverse W-C hydrogen bonding, refers to purine-pyrimidine base pairing, e.g., adenine:thymine, guanine:cytosine, and uracil:adenine. Hoogsteen hydrogen bonding, including reverse Hoogsteen hydrogen bonding, refers to a variation of base pairing in nucleic acids wherein two nucleobases, one on each strand, are held together by hydrogen bonds in the major groove. This non-W-C hydrogen bonding can allow a third strand to wind around a duplex and form triple-stranded helices. Wobble hydrogen bonding, including reverse wobble hydrogen bonding, refers to a pairing between two nucleotides in RNA molecules that does not follow Watson-Crick base pair rules. There are four major wobble base pairs: guanine:uracil, inosine (hypoxanthine):uracil, inosine-adenine, and inosine-cytosine. Rules for canonical hydrogen bonding and non-canonical hydrogen bonding are known to those of ordinary skill in the art (see, e.g., The RNA World, Third Edition (Cold Spring Harbor Monograph Series), R.F. Gesteland, Cold Spring Harbor Laboratory Press (2005), ISBN 978-0879697396; The RNA World, Second Edition (Cold Spring Harbor Monograph Series), R.F. Gesteland, et al., Cold Spring Harbor Laboratory Press (1999), ISBN 978-0879695613; The RNA World (Cold Spring Harbor Monograph Series), R.F. Gesteland, et al., Cold Spring Harbor Laboratory Press (1993), ISBN 978-0879694562 (see, e.g., Appendix 1: Structures of Base Pairs Involving at Least Two Hydrogen Bonds, I. Tinoco); Principles of Nucleic Acid Structure, W. Saenger, Springer International Publishing AG (1988), ISBN 978-0-387-90761-1; Principles of Nucleic Acid Structure, First Edition, S. Neidle, Academic Press (2007), ISBN 978-01236950791).

[0087] "Connect," "connected," and "connecting" are used interchangeably herein and refer to a covalent bond or a non-covalent bond between two macromolecules (e.g., polynucleotides, proteins, and the like).

[0088] As used herein, the terms "nucleic acid sequence," "nucleotide sequence," and "oligonucleotide" are interchangeable and refer to a polymeric form of nucleotides. As used herein, the term "polynucleotide" refers to a polymeric form of nucleotides that has one 5' end and one 3' end, and can comprise one or more nucleic acid sequences. A "circular polynucleotide" refers to a polynucleotide having a covalent bond between its 5' end and its 3' end, thus forming the circular polynucleotide. The nucleotides may be deoxyribonucleotides (DNA), ribonucleotides (RNA), analogs thereof, or combinations thereof (e.g., as described above in the context of guide polynucleotides), and may be of any length. Polynucleotides may perform any function and may have various secondary and tertiary structures. The terms encompass known analogs of natural nucleotides and nucleotides that are modified in the base, sugar, and/or phosphate moieties. Analogs of a particular nucleotide have the same base-pairing specificity (e.g., an analog of A base pairs with T). A polynucleotide may comprise one modified nucleotide or multiple modified nucleotides. Examples of modified nucleotides include, but are not limited to, fluorinated nucleotides, methylated nucleotides, and nucleotide analogs. Nucleotide structure may be modified before or after a polymer is assembled. Following polymerization, polynucleotides may be additionally modified via, for example, conjugation with a labeling component or target binding component. A nucleotide sequence may incorporate non-nucleotide components. Also encompassed are nucleic acids comprising modified backbone residues or linkages, that are synthetic, naturally occurring, and/or non-naturally occurring, and have similar binding properties as a reference polynucleotide (e.g., DNA or RNA). Examples of such analogs include, but are not limited to, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs), Locked Nucleic Acid (LNA™) (Exiqon, Inc., Woburn, MA) nucleosides, glycol nucleic acid, bridged nucleic acids, and morpholino structures.

[0089] Peptide-nucleic acids (PNAs) are synthetic homologs of nucleic acids wherein the polynucleotide phosphate-sugar backbone is replaced by a flexible pseudo-peptide polymer, and nucleobases are linked to the polymer. PNAs have the capacity to hybridize with high affinity and specificity to complementary sequences of RNA and DNA.

[0090] In phosphorothioate nucleic acids, the phosphorothioate (PS) bond replaces a sulfur atom with a non-bridging oxygen in the polynucleotide phosphate backbone. This modification makes the internucleotide linkage resistant to nuclease degradation. In some embodiments, phosphorothioate bonds are introduced between the last 3 to 5 nucleotides at the 5'-end or the 3'-end of a polynucleotide sequence to inhibit exonuclease degradation.

Placement of phosphorothioate bonds throughout an entire oligonucleotide helps reduce degradation by endonucleases, as well.

[0091] Threose nucleic acid (TNA) is an artificial genetic polymer. The backbone structure of TNA comprises repeating threose sugars linked by phosphodiester bonds. TNA polymers are resistant to nuclease degradation. TNA can self-assemble by base-pair hydrogen bonding into duplex structures.

[0092] Linkage inversions can be introduced into polynucleotides through use of "reversed phosphoramidites" (see, e.g., www.ucalgary.ca/dnalab/synthesis/-modifications/linkages). A 3'-3' linkage at a terminus of a polynucleotide stabilizes the polynucleotide to exonuclease degradation by creating an oligonucleotide having two 5'-OH termini but lacking a 3'-OH terminus. Typically, such polynucleotides have phosphoramidite groups on the 5'-OH position and a dimethoxytrityl (DMT) protecting group on the 3'-OH position. Normally, the DMT protecting group is on the 5'-OH and the phosphoramidite is on the 3'-OH.

[0093] Polynucleotide sequences are displayed herein in the conventional 5' to 3' orientation unless otherwise indicated.

[0094] As used herein, "sequence identity" generally refers to the percent identity of nucleotide bases or amino acids comparing a first polynucleotide or polypeptide to a second polynucleotide or polypeptide using algorithms having various weighting parameters. Sequence identity between two polynucleotides or two polypeptides can be determined using sequence alignment by various methods and computer programs (e.g., BLAST, CS-BLAST, PSI-BLAST, FASTA, HMMER, L-ALIGN, and the like) available through the worldwide web at sites including, but not limited to, GENBANK (www.ncbi.nlm.nih.gov/genbank/) and EMBL-EBI (www.ebi.ac.uk). Sequence identity between two polynucleotides or two polypeptide sequences is generally calculated using the standard default parameters of the various methods or computer programs. A high degree of sequence identity, as used herein, between two polynucleotides or two polypeptides is typically between about 90% identity and 100% identity, for example, about 90% identity or higher, preferably about 95% identity or higher, more preferably about 98% identity or higher. A moderate degree of sequence identity, as used herein, between two polynucleotides or two polypeptides is typically between about 80% identity to about 85% identity, for example, about 80% identity or higher, preferably about 85% identity. A low degree of sequence identity, as used herein, between two polynucleotides or two polypeptides is typically between about 50% identity and 75% identity, for example, about 50% identity, preferably about 60% identity, more preferably about 75% identity. For example, a

Cas protein (e.g., Type I-E Cse2, Cas5, Cas6, Cas7, and/or Cas8) comprising amino acid substitutions can have a low degree of sequence identity, a moderate degree of sequence identity, or a high degree of sequence identity over its length to a reference Cas protein (e.g., wild-type Type I-E Cse2, Cas5, Cas6, Cas7, and/or Cas8, respectively). As another example, a guide polynucleotide can have a low degree of sequence identity, a moderate degree of sequence identity, or a high degree of sequence identity over its length compared with a reference wild-type guide polynucleotide that complexes with the reference Cas proteins (e.g., a guide polynucleotide that forms a complex with a Type I-E Cse2, Cas5, Cas6, Cas7, and/or Cas8).

[0095] As used herein, "hybridization" "hybridize," or "hybridizing" is the process of combining two complementary single-stranded DNA or RNA molecules so as to form a single double-stranded molecule (DNA/DNA, DNA/RNA, RNA/RNA) through hydrogen base pairing. Hybridization stringency is typically determined by the hybridization temperature and the salt concentration of the hybridization buffer; e.g., high temperature and low salt provide high stringency hybridization conditions. Examples of salt concentration ranges and temperature ranges for different hybridization conditions are as follows: high stringency, approximately 0.01M to approximately 0.05M salt, hybridization temperature 5°C to 10°C below T_m; moderate stringency, approximately 0.16M to approximately 0.33M salt, hybridization temperature 20°C to 29°C below T_m; and low stringency, approximately 0.33M to approximately 0.82M salt, hybridization temperature 40 °C to 48 °C below T_m. T_m of duplex nucleic acid sequences is calculated by standard methods well known in the art (see, e.g., Maniatis, T., et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press: New York (1982); Casey, J., et al., Nucleic Acids Res. 4:1539-1552 (1977); Bodkin, D.K., et al., J. Virological Methods 10:45-52 (1985); Wallace, R.B., et al., Nucleic Acids Res. 9:879-894 (1981)). Algorithm prediction tools to estimate T_m are also widely available. High stringency conditions for hybridization typically refer to conditions under which a polynucleotide complementary to a target sequence predominantly hybridizes with the target sequence and substantially does not hybridize to non-target sequences. Typically, hybridization conditions are of moderate stringency, preferably high stringency.

[0096] As used herein, "complementarity" refers to the ability of a nucleic acid sequence to form hydrogen bond(s) with another nucleic acid sequence (e.g., through canonical Watson-Crick base pairing). A percent complementarity indicates the percentage of residues in a nucleic acid sequence that can form hydrogen bonds with a second nucleic acid sequence. If two nucleic acid sequences have 100% complementarity, the two sequences are perfectly

complementary, i.e., all of the contiguous residues of a first polynucleotide hydrogen bond with the same number of contiguous residues in a second polynucleotide.

[0097] As used herein, "binding" refers to a non-covalent interaction between macromolecules (e.g., between a protein and a polynucleotide, between a polynucleotide and a polynucleotide, between a protein and a protein, and the like). Such non-covalent interaction is also referred to as "associating" or "interacting" (e.g., if a first macromolecule interacts with a second macromolecule, the first macromolecule binds to the second macromolecule in a non-covalent manner). Some portions of a binding interaction may be sequence-specific (the terms "sequence-specific binding," "sequence-specifically bind," "site-specific binding," and "site specifically binds" are used interchangeably herein). Sequence-specific binding, as used herein, typically refers to one or more guide polynucleotides capable of forming a complex with Type I CRISPR-Cas subunit proteins (e.g., Cse2, Cas5, Cas6, Cas7, and Cas8) to cause the protein to bind a nucleic acid sequence (e.g., a DNA sequence) comprising a nucleic acid target sequence (e.g., a DNA target sequence) preferentially relative to a second nucleic acid sequence (e.g., a second DNA sequence) without the nucleic acid target binding sequence (e.g., the DNA target binding sequence). All components of a binding interaction do not need to be sequence-specific, such as contacts of a protein with phosphate residues in a DNA backbone. Binding interactions can be characterized by a dissociation constant (Kd). "Binding affinity" refers to the strength of the binding interaction. An increased binding affinity is correlated with a lower Kd.

[0098] As used herein, effector complexes are said to "target" a polynucleotide if such a complex binds or cleaves a polynucleotide in the nucleic acid target sequence within the polynucleotide.

[0099] As used herein, a "double-strand break" (DSB) refers to both strands of a double-stranded segment of DNA being severed. In some instances, if such a break occurs, one strand can be said to have a "sticky end" wherein nucleotides are exposed and not hydrogen bonded to nucleotides on the other strand. In other instances, a "blunt end" can occur wherein both strands remain fully base paired with each other.

[00100] "Donor polynucleotide," "donor oligonucleotide," and "donor template" are used interchangeably herein and can be a double-stranded polynucleotide (e.g., DNA), a single-stranded polynucleotide (e.g., DNA or RNA), or a combination thereof. Donor polynucleotides can comprise homology arms flanking the insertion sequence (e.g., DSBs in the DNA). The homology arms on each side can vary in length (e.g., 1-50 bases, 50-100 bases, 100-200 bases, 200-300 bases, 300-500 bases, 500-1000 bases). Homology arms can be symmetric or

asymmetric in length. Parameters for the design and construction of donor polynucleotides are well known in the art (see, e.g., Ran, F., et al., Nature Protocols 8:2281-2308 (2013); Smithies, O., et al., Nature 317:230-234 (1985); Thomas, K., et al., Cell 44:419-428 (1986); Wu, S., et al., Nature Protocols 3:1056-1076 (2008); Singer, B., et al., Cell 31:25-33 (1982); Shen, P., et al., Genetics 112:441-457 (1986); Watt, V., et al., Proc. Natl. Acad. Sci. USA 82:4768-4772 (1985); Sugawara, N., et al., J. Mol. Bio. 12:563-575 (1992); Rubnitz, J., et al., J. Mol. Bio. 4:2253-2258 (1984); Ayares, D., et al., Proc. Natl. Acad. Sci. USA 83:5199-5203 (1986); Liskay, R., et al., Genetics 115:161-167 (1987)). In some embodiments, a donor polynucleotide comprises a chimeric antigen receptor (e.g., a CAR).

[00101] The terms "chimeric antigen receptor" and "CAR" are used interchangeably herein and refer a polypeptide molecule created in the laboratory typically comprising at least two components: an extracellular antigen-recognizing domain (also referred to as a target-binding domain or extracellular ligand binding domain) and an intracellular activation domain (e.g., comprising one or more intracellular signaling domain and typically one or more co-stimulatory signaling domain). A CAR can further comprise a hinge domain and a transmembrane domain. The structure of a typical CAR polypeptide is as follows: N terminus-extracellular–[antigen-recognizing domain–hinge domain]–transmembrane–[transmembrane domain]–intracellular–[intracellular activation domain]-C terminus; or N terminus-intracellular–[intracellular activation domain]-transmembrane–[transmembrane domain]–extracellular–[antigen-recognizing domain–hinge domain]-C terminus.

[00102] Examples of extracellular antigen-recognizing domains comprise moieties used to bind to antigen and include, but are not limited to, single-chain immunoglobulin variable fragment (scFv), an antigen-binding fragment (Fab; typically a region of an antibody that binds an antigen and is composed of one constant and one variable domain of each of the heavy and the light chains), nanobodies, Camelidae family- or shark-derived single chain antibodies, engineered protein binding scaffolds (e.g., DARPins and Centyrins), or natural ligand(s) that bind to their cognate receptor(s).

[00103] Examples of hinge domains include, but are not limited to, a polypeptide hinge of variable length (e.g., one or more amino acids), a hinge region of CD8 alpha, a hinge region of CD28, a hinge region of IgG4, and combinations thereof.

[00104] Examples of transmembrane domains include, but are not limited to, a transmembrane region derived from a transmembrane protein, such as, CD8 alpha, CD28, DAP10, DAP12, NKG2D, and combinations thereof.

[00105] Examples of intracellular activation domains include, but are not limited to, an intracellular signaling domain of CD28, 4-1BB, CD3 zeta, OX40, 2B4, DAP10, DAP12, truncated and mutated signaling domains (e.g., mutations and truncations in the three ITAM domains of CD3 zeta), or other intracellular signaling domains, and combinations thereof.

[00106] When the extracellular ligand binding domain binds to a cognate ligand, the intracellular signaling domain of the CAR activates the lymphocyte (for description of CAR-T cells, see, e.g., Brudno, J., et al., Nature Rev. Clin. Oncol. 15:31-46 (2018); Maude, S., et al., N. Engl. J. Med. 371:1507-1517 (2014); Sadelain, M., et al., Cancer Disc. 3:388-398 (2013); U.S. Patent No. 7,446,190; U.S. Patent No. 8,399,645) (for descriptions of CAR-NK cells, see, e.g., Rezvani, K., et al., Mol. Ther., 25:1769-1781 (2017); Siegler, E., et al., Cell Stem Cell. 23:160-161 (2018); Li, Y., et al., Cell Stem Cell. 23:181-192 (2018); Lin, C., et al., Biochim. Biophys. Acta. Rev. Cancer. 1869:200-215 (2018); Hu, Y., et al., Acta. Pharmacol. Sin. 39:167-176 (2018); Fang, F., et al., Semin. Immunol. 31:37-54 (2017); Glienke, W., et al., Front Pharmacol. 6:21 (2015)).

[00107] Table 2 presents exemplary cellular targets and scFvs/binding proteins that bind the cellular targets. Such scFvs/binding proteins or portions thereof can be incorporated into CAR constructs.

[00108]

Table 2
Exemplary Cellular Targets and CAR scFv Binding Proteins

Cellular target	**CAR scFv/binding protein**
CD19	anti-CD19
CD20	anti-CD20
CD22	anti-CD22
CD30	anti-CD30
CD33	anti-CD33
CD37	anti-CD37
CD43	anti-CD43
CD138	anti-CD138
CD171/L1CAM	anti-CD171
CEA	anti-CEA

Table 2 Exemplary Cellular Targets and CAR scFv Binding Proteins	
Cellular target	**CAR scFv/binding protein**
CD123	anti-CD123
B-cell activating factor receptor (BAFF-R) [also called, Tumor necrosis factor receptor superfamily member 13C (TNFRSF13C); BLyS receptor 3 (BR3); and CD268	Anti-BAFF-R
IL13 Receptor alpha	IL13
Epidermal growth factor receptor	anti-Epidermal growth factor receptor
EFGRvIII	anti-EFGRvIII
ErbB	anti-ErbB
FAP	anti-FAP
GD2	anti-GD2
Glypican 3	anti-Glypican 3
Her2	anti-Her2
Mesothelin	anti-Mesothelin
ULBP and MICA/B proteins	NKG2D
PD1	anti-PD1
MUC1	anti-MUC1
VEGF2	anti-VEGF2
SLAMF7	anti-SLAMF7
BCMA	anti-BCMA
WT1	anti-WT1
MUC16	anti-MUC16
LewisY/LeY	anti-LeY
FLT3	FLT3 ligand or anti-FLT3
ROR1	anti-ROR1
Claudin18	Anti-Claudin18

Table 2 Exemplary Cellular Targets and CAR scFv Binding Proteins	
Cellular target	**CAR scFv/binding protein**
Claudin6	Anti-Claudin6

[00109] As used herein, "homology-directed repair" (HDR) refers to DNA repair that takes place in cells, for example, during repair of a DSB in gDNA. HDR requires nucleotide sequence homology and uses a donor or template polynucleotide to repair the sequence wherein the DSB (e.g., within a DNA target sequence) occurred. The donor polynucleotide generally has the requisite sequence homology with the sequence flanking the DSB so that the donor polynucleotide can serve as a suitable template for repair. HDR results in the transfer of genetic information from, for example, the donor polynucleotide to the DNA target sequence. HDR may result in alteration of the DNA target sequence (e.g., insertion, deletion, or mutation) if the donor polynucleotide sequence differs from the DNA target sequence and part or all of the donor polynucleotide is incorporated into the DNA target sequence. In some embodiments, an entire donor polynucleotide, a portion of the donor polynucleotide, or a copy of the donor polynucleotide is integrated at the site of the DNA target sequence. For example, a donor polynucleotide can be used for repair of the break in the DNA target sequence, wherein the repair results in the transfer of genetic information from the donor polynucleotide at the site or in close proximity of the break in the DNA. Accordingly, new genetic information may be inserted or copied at a DNA target sequence.

[00110] A "genomic region" is a segment of a chromosome in the genome of a host cell that is present on either side of the nucleic acid target sequence site or, alternatively, also includes a portion of the nucleic acid target sequence site. The homology arms of the donor polynucleotide have sufficient homology to undergo homologous recombination with the corresponding genomic regions. In some embodiments, the homology arms of the donor polynucleotide share significant sequence homology to the genomic region immediately flanking the nucleic acid target sequence site; it is recognized that the homology arms can be designed to have sufficient homology to genomic regions farther from the nucleic acid target sequence site.

[00111] As used herein, "non-homologous end joining" (NHEJ) refers to the repair of a DSB in DNA by direct ligation of one terminus of the break to the other terminus of the break without a requirement for a donor polynucleotide. NHEJ is a DNA repair pathway available to cells to repair DNA without the use of a repair template. NHEJ in the absence of a donor

polynucleotide often results in nucleotides being randomly inserted or deleted at the site of the DSB.

[00112] "Microhomology-mediated end joining" (MMEJ) is pathway for repairing a DSB in gDNA. MMEJ involves deletions flanking a DSB and alignment of microhomologous sequences internal to the break site before joining. MMEJ is genetically defined and requires the activity of, for example, CtIP, Poly(ADP-Ribose) Polymerase 1 (PARP1), DNA polymerase theta (Pol θ), DNA Ligase 1 (Lig 1), or DNA Ligase 3 (Lig 3). Additional genetic components are known in the art (see, e.g., Sfeir, A., et al., Trends in Biochemical Sciences 40:701-714 (2015)).

[00113] As used herein, "DNA repair" encompasses any process whereby cellular machinery repairs damage to a DNA molecule contained in the cell. The damage repaired can include single-strand-breaks or DSBs. At least three mechanisms exist to repair DSBs: HDR, NHEJ, and MMEJ. "DNA repair" is also used herein to refer to DNA repair resulting from human or machine manipulation, wherein a target locus is modified, e.g., by inserting, deleting, or substituting nucleotides, all of which represent forms of genome editing.

[00114] As used herein, "recombination" refers to a process of exchange of genetic information between two polynucleotides.

[00115] As used herein, the terms "regulatory sequences," "regulatory elements," and "control elements" are interchangeable and refer to polynucleotide sequences that are upstream (5' non-coding sequences), within, or downstream (3' non-translated sequences) of a polynucleotide target to be expressed. Regulatory sequences influence, for example, the timing of transcription; the amount or level of transcription; RNA processing or stability; and/or translation of the related structural nucleotide sequence. Regulatory sequences may include activator binding sequences, enhancers, introns, polyadenylation recognition sequences, promoters, transcription start sites, repressor binding sequences, stem-loop structures, translational initiation sequences, internal ribosome entry sites (IRES), translation leader sequences, transcription termination sequences (e.g., polyadenylation signals and poly-U sequences), translation termination sequences, primer binding sites, and the like.

[00116] Regulatory elements include those that direct constitutive, inducible, and repressible expression of a nucleotide sequence in many types of host cells and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). In some embodiments, a vector comprises one or more pol III promoters, one or more pol II promoters, one or more pol I promoters, or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II

promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer; see, e.g., Boshart, M., et al., Cell 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter, as well as engineered artificial promoters (e.g., the MND promoter and the CAG promoter). It will be appreciated by those skilled in the art that the design of an expression vector may depend on such factors as the choice of the host cell to be transformed, the level of expression desired, and the like. A vector can be introduced into host cells to thereby produce RNA transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acid sequences as described herein.

[00117] "Gene," as used herein, refers to a polynucleotide sequence comprising exon(s) and related regulatory sequences. A gene may further comprise intron(s) and/or untranslated region(s) (UTR(s)).

[00118] As used herein, the term "operably linked" refers to polynucleotide sequences or amino acid sequences placed into a functional relationship with one another. For example, regulatory sequences (e.g., a promoter or enhancer) are "operably linked" to a polynucleotide encoding a gene product if the regulatory sequences regulate or contribute to the modulation of the transcription of the polynucleotide. Operably linked regulatory elements are typically contiguous with the coding sequence. However, enhancers can function if separated from a promoter by up to several kilobases or more. Additionally, multicistronic constructs can include multiple coding sequences that use only one promoter by including a 2A self-cleaving peptide, an IRES element, etc. Accordingly, some regulatory elements may be operably linked to a polynucleotide sequence but not contiguous with the polynucleotide sequence. Similarly, translational regulatory elements contribute to the modulation of protein expression from a polynucleotide.

[00119] As used herein, "expression" refers to transcription of a polynucleotide from a DNA template, resulting in, for example, a messenger RNA (mRNA) or other RNA transcript (e.g., non-coding, such as structural or scaffolding RNAs). The term further refers to the process through which transcribed mRNA is translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be referred to collectively as "gene product(s)." Expression may include splicing the mRNA in a eukaryotic cell, if the polynucleotide is derived from gDNA.

[00120] A "coding sequence" or a sequence that "encodes" a selected polypeptide, is a nucleic acid molecule that is transcribed (in the case of DNA) and translated (in the case of

mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' terminus and a translation stop codon at the 3' terminus.

[00121] By "artificial transcriptional activator (ATA)" or an "artificial transcription factor (ATF)," as used herein, is meant a complex capable of recruiting RNA polymerase II holoenzyme to genes with which they are associated thereby causing ectopic expression of the gene of interest. Such activators include at least two components: (1) a catalytically inactive polynucleotide binding domain that either directly recognizes cognate nucleotide sequences and can bind to these sequences, or a polynucleotide binding domain that is guided to such sequences for binding (e.g., a nucleoprotein complex comprising a nucleic acid binding domain and a guide as described herein); and (2) an activation domain (also termed "effector domain") that interacts with a variety of proteins that constitute the transcriptional machinery to upregulate transcription.

[00122] By "catalytically inactive polynucleotide binding domain" is meant a molecule that binds to, but does not cleave, the nucleic acid target site bound by the binding domain. Representative examples of such domains are detailed herein.

[00123] As used herein, the term "modulate" refers to a change in the quantity, degree, or amount of a function. For example, a Type I CRISPR nucleoprotein complex, as disclosed herein, may modulate the activity of a promoter sequence by binding to a nucleic acid target sequence at or near the promoter or a transcriptional start site or regulator site. Depending on the action occurring after binding, the Type I CRISPR nucleoprotein complex can induce, enhance, suppress, or inhibit transcription of a gene operatively linked to the promoter sequence. Thus, "modulation" of gene expression includes both gene activation and gene repression.

[00124] Modulation can be assayed by determining any characteristic directly or indirectly affected by the expression of the target gene. Such characteristics include, for example, changes in RNA or protein levels, protein activity, product levels, expression of the gene, or activity level of reporter genes. Accordingly, the terms "modulating expression," "inhibiting expression," and "activating expression" of a gene can refer to the ability of a Type I CRISPR nucleoprotein complex to change, activate, or inhibit transcription of a gene.

[00125] A function (e.g., an enzymatic function) can be up-modulated (e.g., increase, strengthen, amplify, or enhance the function) or down-modulated (e.g., decrease, weaken, diminish, or lessen the function). In one embodiment, binding of a mCas3 protein to single-

stranded DNA (ssDNA) or ATP binding/hydrolysis by a mCas3 protein can be up-modulated or down-modulated relative to the corresponding wtCas3 protein.

[00126] "Vector" and "plasmid," as used herein, refer to a polynucleotide vehicle to introduce genetic material into a cell. Vectors can be linear or circular. Vectors can contain a replication sequence capable of effecting replication of the vector in a suitable host cell (e.g., an origin of replication). Upon transformation of a suitable host, the vector can replicate and function independently of the host genome or integrate into the host genome. Vector design depends, among other things, on the intended use and host cell for the vector, and the design of a vector of the invention for a particular use and host cell is within the level of skill in the art. The four major types of vectors are plasmids, viral vectors, cosmids, and artificial chromosomes. Typically, vectors comprise an origin of replication, a multicloning site, and/or a selectable marker. An expression vector typically comprises an expression cassette. By "recombinant virus" is meant a virus that has been genetically altered, e.g., by the addition or insertion of a heterologous nucleic acid construct into a viral genome or portion thereof.

[00127] As used herein, "expression cassette" refers to a polynucleotide construct generated using recombinant methods or by synthetic means and comprising regulatory sequences operably linked to a selected polynucleotide to facilitate expression of the selected polynucleotide in a host cell. For example, the regulatory sequences can facilitate transcription of the selected polynucleotide in a host cell, or transcription and translation of the selected polynucleotide in a host cell. An expression cassette can, for example, be integrated in the genome of a host cell or be present in a vector to form an expression vector.

[00128] As used herein, a "targeting vector" is a recombinant DNA construct typically comprising tailored DNA arms, homologous to gDNA, that flank elements of a target gene or nucleic acid target sequence (e.g., a DSB). A targeting vector comprises a donor polynucleotide. Elements of the target gene can be modified in a number of ways, including deletions and/or insertions. A defective target gene can be replaced by a functional target gene, or in the alternative a functional gene can be knocked out. Optionally, the donor polynucleotide of a targeting vector comprises a selection cassette comprising a selectable marker that is introduced into the target gene. Targeting regions (comprising nucleic acid target sequences) adjacent or within a target gene can be used to affect regulation of gene expression.

[00129] As used herein, the term "between" is inclusive of end values in a given range (e.g., between 1 and 50 nucleotides in length includes 1 nucleotide and 50 nucleotides; between 5 amino acids and 50 amino acids in length includes 5 amino acids and 50 amino acids).

[00130] As used herein, the term "amino acid" (aa) refers to natural and synthetic (unnatural) amino acids, including amino acid analogs, modified amino acids, peptidomimetics, glycine, and D or L optical isomers.

[00131] As used herein, the terms "peptide," "polypeptide," "protein," and "subunit protein" are interchangeable and refer to polymers of amino acids. A polypeptide may be of any length. It may be branched or linear, it may be interrupted by non-amino acids, and it may comprise modified amino acids. The terms also refer to an amino acid polymer that has been modified through, for example, acetylation, disulfide bond formation, glycosylation, lipidation, phosphorylation, pegylation, biotinylation, cross-linking, and/or conjugation (e.g., with a labeling component or ligand). Polypeptide sequences are displayed herein in the conventional N-terminal to C-terminal orientation, unless otherwise indicated.

[00132] Polypeptides and polynucleotides can be made using routine techniques in the field of molecular biology (see, e.g., standard texts listed above). Furthermore, essentially any polypeptide or polynucleotide is available from commercial sources.

[00133] The terms "fusion protein" and "chimeric protein," as used herein, refer to a single protein created by joining two or more proteins, protein domains, protein fragments, or circular permuted polypeptides that do not naturally occur together in a single protein. In some embodiments, a linker polynucleotide can be used to connect a first protein, protein domains, or protein fragments, or circular permuted polypeptides to a second protein, protein domains, protein fragments, or circular permuted polypeptides. For example, a fusion protein can comprise a Type I CRISPR-Cas protein (e.g., Cas8, Cas3) and a functional domain from another protein (e.g., FokI; see, e.g., U.S. Patent No. 9,885,026). The modification to include such domains in fusion proteins may confer additional activity on engineered Type I CRISPR-Cas proteins. Such activities can include nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity, glycosylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, and/or myristoylation activity or demyristoylation activity that modifies a polypeptide associated with nucleic acid target sequence (e.g., a histone).

[00134] In some embodiments, a fusion protein can comprise epitope tags (e.g., histidine tags, HA tags, FLAG® (Sigma Aldrich, St. Louis, MO) tags, Myc tags, nuclear localization signal (NLS) tags, SunTag), reporter protein sequences (e.g., glutathione-S-transferase, beta-galactosidase, luciferase, green fluorescent protein, cyan fluorescent protein, yellow fluorescent protein), and/or nucleic acid sequence binding domains (e.g., a DNA binding domain or an RNA binding domain).

[00135] A fusion protein can also comprise activator domains (e.g., heat shock transcription factors, NFKB activators) or repressor domains (e.g., a KRAB domain). As described by Lupo, A., et al., Current Genomics 14:268-278 (2013), the KRAB domain is a potent transcriptional repression module and is located in the amino-terminal sequence of most C2H2 zinc finger proteins (see, e.g., Margolin, J., et al., Proc. Natl. Acad. Sci. USA 91:4509-4513 (1994); Witzgall, R., et al., Proc. Natl. Acad. Sci. USA 91:4514-4518 (1994)). The KRAB domain typically binds to co-repressor proteins and/or transcription factors via protein-protein interactions, causing transcriptional repression of genes to which KRAB zinc finger proteins (KRAB-ZFPs) bind (see, e.g., Friedman, J.R., et al., Genes & Development 10:2067-2678 (1996)). In some embodiments, linker nucleic acid sequences are used to join the two or more proteins, protein domains, or protein fragments.

[00136] As used herein, "CASCADEa" (Cascade activation) is a CRISPR method or system wherein the method or system activates the expression of a gene associated with the locus of the target nucleic acid sequence of a Cascade RNP complex. In some embodiments, one or more proteins of a Cascade complex are fused to an effector domain (e.g., VP16 or VP64) and a Cascade RNP complex comprising the fusion and guide polynucleotide is used for the recruitment of endogenous transcription factors. In some embodiments, the guide polynucleotide can be fused 5' or 3' to a nucleotide effector domain such as an MS2 binding RNA that also recruits transcription factors.

[00137] As used herein, "CASCADEi" (Cascade inhibition) is a CRISPR method or system wherein the CRISPR method or system down-regulates the expression of a gene associated with the locus of the target nucleic acid sequence of a Cascade RNP complex (i.e., a Cascade RNP complex is used down-regulate the expression of the gene). For the recruitment of endogenous repression factors, one or more proteins in a Cascade complex is typically fused to an effector domain (e.g., KRAB). In some embodiments, the guide polynucleotide can be fused 5' or 3' to a nucleotide effector domain that also recruits endogenous transcriptional repression effector proteins.

[00138] A "moiety," as used herein, refers to a portion of a molecule. A moiety can be a functional group or describe a portion of a molecule with multiple functional groups (e.g., that share common structural aspects). The terms "moiety" and "functional group" are typically used interchangeably herein; however, a "functional group" can more specifically refer to a portion of a molecule that comprises some common chemical behavior. "Moiety" is often used as a structural description. In some embodiments, a 5' terminus, a 3' terminus, or a 5' terminus and a 3' terminus (e.g., a non-native 5' terminus and/or a non-native 3' terminus in a first stem element) can comprise one or more moieties.

[00139] As used herein, "adoptive cell" refers to a cell that can be genetically modified for use in a cell therapy treatment, such for treating cancer and/or preventing graft versus host disease (GvHD) and other undesirable side-effects of cell therapies, such as, but not limited to, cytokine storm, oncogenic transformations of the administered genetically modified material, neurological disorders, and the like. Adoptive cells include, but are not limited to, stem cells, induced pluripotent stem cells (iPSCs), cord blood stem cells, lymphocytes, macrophages, red blood cells, fibroblasts, endothelial cells, epithelial cells, and pancreatic precursor cells.

[00140] As used herein, "cell therapy" refers to the treatment of a disease or disorder that utilizes genetically modified cells. Genetic modifications can be introduced using methods described herein, such as methods comprising viral vectors, nucleofection, gene gun delivery, sonoporation, cell squeezing, lipofection, or the use of other chemicals, cell penetrating peptides, and the like.

[00141] As used herein, "adoptive cell therapy (ACT)" refers to a therapy that uses genetically modified adoptive cells derived from either a specific patient returned to that patient (autologous cell therapy) or from a third-party donor (allogeneic cell therapy), to treat the patient. ACTs, include, but are not limited to, bone marrow transplants, stem cell transplants, T-cell therapies, CAR-T cell therapies, and natural killer (NK) cell therapies.

[00142] As used herein, "lymphocyte" refers to a leukocyte (white blood cell) that is part of the vertebrate immune system. Also encompassed by the term "lymphocyte" is a hematopoietic stem cell or an induced pluripotent stem cells (iPSC) that gives rise to lymphoid cells. Lymphocytes include T cells for cell-mediated, cytotoxic adaptive immunity, such as CD4+ and/or CD8+ cytotoxic T cells; alpha/beta T cells and gamma/delta T cells; regulatory T cells, such as Treg cells; NK cells that function in cell-mediated, cytotoxic innate immunity; B cells, for humoral, antibody-driven adaptive immunity; NK/T cells; cytokine induced killer cells (CIK cells); and antigen presenting cells (APCs), such as dendritic cells. A lymphocyte can be a mammalian cell, such as a human (*Homo sapiens*; *H. sapiens*) cell. The term

"lymphocyte" also encompasses genetically modified T cells and NK cells, modified to produce chimeric antigen receptors (CARs) on the T or NK cell surface (CAR-T cells and CAR-NK cells). These CAR-T cells recognize specific soluble antigens or antigens on a target cell surface, such as a tumor cell surface, or on cells in the tumor microenvironment.

[00143] Also encompassed by the term "lymphocyte," as used herein, are T-cell receptor engineered T cells (TCRs), genetically engineered to express one or more specific, naturally occurring or engineered T-cell receptors that can recognize protein or (glyco)lipid antigens of target cells presented by the Major Histocompatibility Complex (MHC). Small pieces of these antigens, such as peptides or fatty acids, are shuttled to the target cell surface and presented to the T-cell receptors as part of the MHC. T-cell receptor binding to antigen-loaded MHCs activates the lymphocyte.

[00144] Lymphocyte activation occurs when lymphocytes are triggered through antigen-specific receptors on their cell surface. This causes the cells to proliferate and differentiate into specialized effector lymphocytes. Such "activated" lymphocytes are typically characterized by a set of receptors on the surface of the lymphocyte. Surface markers for activated T cells include CD3, CD4, CD8, PD1, IL2R, and others. Activated cytotoxic lymphocytes can kill target cells after binding cognate receptors on the surface of target cells.

[00145] Tumor infiltrating lymphocytes (TILs) are also encompassed by the term "lymphocyte," as used herein. TILs are immune cells that have penetrated the environment in and around a tumor ("the tumor microenvironment"). TILs are typically isolated from tumor cells and the tumor microenvironment and are selected *in vitro* for high reactivity against tumor antigens. TILs are grown *in vitro* under conditions that overcome the tolerizing influences that exist *in vivo* and are then introduced into a subject for treatment.

[00146] T cells typically are present in a number of subtypes such as "naive T cells" (Tn), "Stem cell memory T cells" (Tscm), "Central memory T cells" (Tcm) "Effector memory T cells" (Tem) ,"Effector T cells" (Teff) and "regulatory T cells" (Treg). Each T-cell subset is characterized by a set of cell surface markers.

[00147] The term "affinity tag," as used herein, typically refers to one or more moieties that increases the binding affinity of one macromolecule for another, for example, to facilitate formation of an engineered Type I CRISPR-Cas nucleoprotein complex. In some embodiments, an affinity tag can be used to increase the binding affinity of one Cas subunit protein for another Cas subunit protein (e.g., a first Cas7 protein for a second Cas7 protein). In some embodiments, an affinity tag can be used to increase the binding affinity of one or more Cas subunit proteins for a cognate guide polynucleotide. Some embodiments of the present

invention introduce one or more affinity tags to the N-terminal of a Cas subunit protein sequence, to the C-terminal of a Cas subunit protein sequence, to a position located between the N-terminal and C-terminal of a Cas subunit protein sequence, or to combinations thereof. In some embodiments of the present invention, one or more guide polynucleotide comprises an affinity tag that increases binding affinity of the guide polynucleotide with one or more Cas subunit proteins. A wide variety of affinity tags are disclosed in U.S. Published Patent Application No. 2014-0315985, published 23 October 2014. Ligands and ligand-binding moieties are paired affinity tags.

[00148] As used herein, a "cross-link" is a bond that links one polymer chain (e.g., a polynucleotide or polypeptide) to another. Such bonds can be covalent bonds or ionic bonds. In some embodiments, one polynucleotide can be bound to another polynucleotide by cross linking the polynucleotides. In other embodiments, a polynucleotide can be cross linked to a polypeptide. In additional embodiments, a polypeptide can be cross linked to a polypeptide.

[00149] The term "cross-linking moiety," as used herein, typically refers to a moiety suitable to provide cross linking between two macromolecules. A cross-linking moiety is another example of an affinity tag.

[00150] As used herein, a "host cell" generally refers to a biological cell. A cell is the basic structural, functional, and/or biological unit of an organism. A cell can originate from any organism having one or more cells. Examples of host cells include, but are not limited to, a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a cell of a eukaryotic organism, a protozoal cell, a cell from a plant, an algal cell, (e.g., *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens C. agardh*, and the like), seaweeds (e.g., kelp), a fungal cell (e.g., a yeast cell or a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g., fruit fly, cnidarian, echinoderm, nematode, and the like), a cell from a vertebrate animal including mammals (e.g., a pig, a cow, a goat, a sheep, a rodent, a rat, a mouse, a non-human primate, a human, etc.). Furthermore, a host cell can be a stem cell or progenitor cell, and an immunological cell, such as any of the immunological cells described herein. The host cell can be a human cell. In some embodiments, the human cell is outside of the human body. In some embodiments, cells of a body of a living organism (e.g., a human body) are manipulated *ex vivo* (i.e., outside of the living body). *Ex vivo* often refers to a medical procedure in which an organ, cells, or tissue are taken from a living body (e.g., a human body) for a treatment or procedure, and then returned to the living body.

[00151] As used herein, "stem cell" refers to a cell that has the capacity for self-renewal, i.e., the ability to go through numerous cycles of cell division while maintaining the undifferentiated state. Stem cells can be totipotent, pluripotent, multipotent, oligopotent, or unipotent. Stem cells can be embryonic, fetal, amniotic, adult, or induced pluripotent stem cells.

[00152] As used herein, "induced pluripotent stem cell" refers to a type of pluripotent stem cell that is artificially derived from a non-pluripotent cell, typically a somatic cell. In some embodiments, the somatic cell is a human somatic cell. Examples of somatic cells include, but are not limited to, dermal fibroblasts, bone marrow-derived mesenchymal cells, cardiac muscle cells, keratinocytes, liver cells, stomach cells, neural stem cells, lung cells, kidney cells, spleen cells, and pancreatic cells. Additional examples of somatic cells include cells of the immune system, including but not limited to, B cells, dendritic cells, granulocytes, innate lymphoid cells, megakaryocytes, monocytes/macrophages, myeloid-derived suppressor cells, natural killer (NK) cells, T cells, thymocytes, and hematopoietic stem cells.

[00153] As used herein, "hematopoietic stem cell" refers to an undifferentiated cell that has the ability to differentiate into a hematopoietic cell, such as a lymphocyte.

[00154] "Plant," as used herein, refers to whole plants, plant organs, plant tissues, germplasm, seeds, plant cells, and progeny of the same. Plant cells include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant parts include differentiated and undifferentiated tissues including, but not limited to, roots, stems, shoots, leaves, pollens, seeds, tumor tissue, and various forms of cells and culture (e.g., single cells, protoplasts, embryos, and callus tissue). The plant tissue may be in plant or in a plant organ, tissue, or cell culture. "Plant organ" refers to plant tissue or a group of tissues that constitute a morphologically and functionally distinct part of a plant.

[00155] The terms "subject," "individual," or "patient" are used interchangeably herein and refer to any member of the phylum Chordate, including, without limitation, humans and other primates, including non-human primates such as rhesus macaques, chimpanzees, and other monkey and ape species; farm animals, such as cattle, sheep, pigs, goats, and horses; domestic mammals, such as dogs and cats; laboratory animals, including rabbits, mice, rats, and guinea pigs; birds, including domestic, wild, and game birds, such as chickens, turkeys, and other gallinaceous birds, ducks, and geese; and the like. The term does not denote a particular age or gender. Thus, the term includes adult, young, and newborn individuals as well as males and females. In some embodiments, a host cell is derived from a subject (for example, lymphocytes, stem cells, progenitor cells, or tissue-specific cells). In some embodiments, the

subject is a non-human subject. In some embodiments, the subject is a human (*H. sapiens*) subject.

[00156] The terms "effective amount" or "therapeutically effective amount" of a composition or agent, such as a genetically engineered adoptive cell as provided herein, refer to a sufficient amount of the composition or agent to provide the desired response, such as to prevent or eliminate one or more harmful side-effects associated with allogeneic adoptive cell therapies. Such responses will depend on the particular disease in question. For example, in a patient being treated for cancer using an adoptive cell therapy, a desired response includes, but is not limited to, treatment or prevention of the effects of GvHD, Host versus Graft rejection, cytokine release syndrome (CRS), cytokine storm, and the reduction of oncogenic transformations of administered genetically modified cells. 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 condition being treated, and the particular modified lymphocyte used, mode of administration, and the like. An appropriate "effective" amount in any individual case may be determined by one of ordinary skill in the art using routine experimentation.

[00157] "Treatment" or "treating" a particular disease, such as cancerous condition, or GvHD, includes: (1) preventing the disease, for example, preventing the development of the disease or causing the disease to occur with less intensity in a subject that may be predisposed to the disease but does not yet experience or display symptoms of the disease; (2) inhibiting the disease, for example, reducing the rate of development, arresting the development or reversing the disease state; and/or (3) relieving symptoms of the disease, for example, decreasing the number of symptoms experienced by the subject.

[00158] By "gene editing" or "genome editing," as used herein, is meant a type of genetic engineering that results in a genetic modification, such as an insertion, deletion, or replacement of a nucleotide sequence, or even a single base, at a specific site in a cell genome. The terms include, without limitation, heterologous gene expression, gene or promoter insertion or deletion, nucleic acid mutation, and a disruptive genetic modification, as defined herein.

[00159] By "epitope" is meant a site on a molecule to which specific B cells and T cells respond. An epitope can comprise 3 or more amino acids in a spatial conformation unique to the epitope. Generally, an epitope consists of at least five such amino acids and, more usually, consists of at least 8-10 such amino acids. Methods of determining spatial conformation of amino acids are known in the art and include, for example, x-ray crystallography, electron microscopy, and 2-dimensional nuclear magnetic resonance. Furthermore, the identification of

epitopes in a given protein is readily accomplished using techniques well known in the art, such as by the use of hydrophobicity studies and by site-directed serology.

[00160] A "mimotope" is a macromolecule, such as a peptide, that mimics the structure of an epitope. Because of this property, it causes an antibody response similar to the one elicited by the epitope. An antibody for a given epitope antigen will recognize a mimotope that mimics that epitope. Mimotopes are commonly obtained from phage display libraries through biopanning.

[00161] An "antibody" intends a molecule that "recognizes," i.e., specifically binds to an epitope of interest present in a polypeptide, such as a ligand binding domain. By "specifically binds" is meant that the antibody interacts with the epitope in a "lock and key" type of interaction to form a complex between the antigen and antibody. The term "antibody," as used herein, includes antibodies obtained from monoclonal preparations, as well as, the following: hybrid (chimeric) antibody molecules; F(ab')2 and F(ab) fragments; Fv molecules (non-covalent heterodimers; single-chain Fv molecules (scFv); dimeric and trimeric antibody fragment constructs; minibodies; humanized antibody molecules; single chain antibodies; Nanobody® (Ablynx N.V., Zwijnaarde, Belgium) antibodies; and any functional fragments obtained from such molecules, wherein such fragments retain immunological binding properties of the parent antibody molecule. The antibodies can be sourced from different species, such as human, mouse, rat, rabbit, camel, chicken, and the like. Antibodies and antibody parts can then be further obtained by *in vitro* techniques, such as by phage display and yeast display. Fully humanized antibodies can be obtained from human plasma, human B cell cloning, mouse, rat, rabbit, chicken, etc., that have an engineered humanized B cell repertoire. Antibodies can then be further modified by affinity maturation and other methods, such as afucosylation or IgG Fc engineering.

[00162] As used herein, the term "monoclonal antibody" refers to an antibody composition having a homogeneous antibody population. The term is not limited regarding the species or source of the antibody, nor is it intended to be limited by the manner in which it is made. The term encompasses whole immunoglobulins as well as fragments such as Fab, $F(ab')_2$, Fv, and other fragments, as well as chimeric and humanized homogeneous antibody populations, that exhibit immunological binding properties of the parent monoclonal antibody molecule.

[00163] "Antibody-dependent cell-mediated cytotoxicity (ADCC)" also referred to as "antibody-dependent cellular cytotoxicity," refers to a mechanism whereby an effector cell of the immune system actively lyses a target cell, such as an adoptive cell, when a membrane-

surface ligand binding domain has been bound by a specific antibody. Effector cells are typically natural killer (NK) cells. However, macrophages, neutrophils, and eosinophils can also mediate ADCC. ADCC is independent of complement-dependent cytotoxicity (CDC) that also lyses targets by damaging membranes without the involvement of antibodies or cells of the immune system.

[00164] "Transformation," as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for insertion. For example, transformation can be by direct uptake, transfection, infection, and the like. The exogenous polynucleotide may be maintained as a nonintegrated vector, for example, an episome, or, alternatively, may be integrated into the host genome. As used herein, "transgenic organism" refers to an organism that contains genetic material into which DNA from an unrelated organism has been artificially introduced. The term includes the progeny (any generation) of a transgenic organism, provided that the progeny has the genetic modification. In some embodiments, the transgenic organism is a non-human transgenic organism.

[00165] As used herein, "isolated" can refer to a molecule (e.g., a polynucleotide or a polypeptide) that, by human intervention, exists apart from its native environment and is therefore not a product of nature. When referring to a polypeptide, isolated means that the indicated molecule is separate and discrete from the whole organism with which the molecule is found in nature or is present in the substantial absence of other biological macromolecules of the same type. The term "isolated" with respect to a polynucleotide is a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but having heterologous sequences in association therewith; or a molecule disassociated from the chromosome.

[00166] The term "purified," as used herein, preferably means at least 75% by weight, more preferably at least 85% by weight, more preferably still at least 95% by weight, and most preferably at least 98% by weight, of the same molecule is present.

[00167] As used herein, a "substrate channel" refers to the direct transfer of a reactant from one enzymatic reaction to another enzymatic reaction without first diffusing into the bulk environment (see, e.g., Wheeldon, I., et al., Nat. Chem. 8:299-309 (2016)). Intermediates of these enzymatic steps are not in equilibrium with the bulk solution, which enables the increased efficiencies and yields in enzymatic processes. Frequently, enzymes in naturally occurring metabolic processes have evolved means of co-localization and assembly into controlled aggregates.

[00168] As used herein, "substrate channel element" refers to a component of a metabolic pathway. In some embodiments, a substrate channel element is an enzyme that catalyzes a chemical reaction.

[00169] As used herein, "substrate channel complex" refers to multiple substrate channel elements that are co-localized together via some means.

[00170] As used herein, an "RNA scaffold" refers to an RNA molecule that peptides can use as a substrate for binding.

[00171] The data presented herein demonstrate that fusions between Cascade components and nuclease domains (e.g., a dimerization-dependent, non-specific FokI nuclease domains; see, e.g., Urnov, F.D., et al., Nature Reviews Genetics 11:636–646 (2010); Joung, J.K., et al., Nat. Rev. Mol. Cell Biol. 14:49–55 (2013); Guilinger, J.P., et al., Nat. Biotechnol. 32:577–582 (2014); Tsai, S.Q., et al., Nat. Biotechnol. 32:569–576 (2014)) mediate efficient programmable RNA-guided gene editing with Type I systems in human cells. The data demonstrate that engineered Type I CRISPR-Cas systems (e.g., a comprising FokI-Cascade component fusion) can be directly transfected as intact ribonucleoprotein (RNP) complexes or assembled in cells via delivery of individual plasmid-encoded components. As set forth herein, all the CRISPR-associated (Cas) genes were assembled onto a single polycistronic vector, yielding a simplified two-component Cas protein-guide RNA expression system. In addition, length/composition design of the nuclease (e.g., FokI)/Cascade component linker sequences and formulation of appropriate DNA geometry, as well as selective Cascade homolog choice, provide engineered Type I CRISPR-Cas complexes having editing efficiencies up to about 50%. Key characteristics of the engineered Type I CRISPR-Cas systems (e.g., comprising FokI–Cascade component fusion proteins) related to PAM requirements and mismatch sensitivities during DNA targeting were determined.

[00172] In a first aspect, the present invention relates to engineered polynucleotides encoding Cascade components including, but not limited to, Cascade subunit proteins and Cascade guide polynucleotides.

[00173] In one embodiment, the present invention relates to engineered polynucleotides encoding Cascade components that are derived from Cascade Type I-E systems. Exemplary polynucleotide constructs comprising Cascade proteins and Cascade crRNAs are presented in Example 1. Example 1, Table 15, and SEQ ID NO:1 through SEQ ID NO:20 present polynucleotide DNA sequences of genes encoding the five subunit proteins of Type I-E Cascade, specifically from *E. coli* strain K-12 MG1655, as well as the amino acid sequences of the resulting protein components. The polynucleotide sequences were derived from *E. coli*

gDNA and were codon-optimized specifically for expression in *E. coli*, and/or codon-optimized specifically for expression in eukaryotic cells (e.g., human cells). When this polynucleotide is transcribed into a precursor crRNA and processed by the Cascade RNA endonuclease, a mature crRNA is produced that functions as a guide RNA to target complementary DNA sequences in the genome. The minimal CRISPR array comprises two repeat sequences (underlined in the CRISPR array sequences presented in Example 1) flanking an exemplary spacer sequence, which represents the guide portion of the crRNA. RNA processing by the Cascade endonuclease generates a crRNA with repeat sequences on both the 5' and 3' ends, flanking the guide sequence. One of ordinary skill in the art, in view of the teachings of the present Specification and the Examples, can select appropriate spacer sequences to target binding of a Cascade complex to a chosen target sequence (e.g., in gDNA).

[00174] Polynucleotide sequences encoding Cascade components from additional bacterial or archaeal species can be identified and designed following the guidance of the present Specification and using bioinformatics tools such as BLAST and PSI-BLAST to locate, as an example, homologs of Cascade subunit genes from *E. coli* strain K-12 MG1655, and then inspecting the flanking genomic neighborhood of the Cascade gene to locate and identify genes of the remaining Cascade subunit proteins (see, e.g., Example 14A, Example 14B, Example 15A, and Example 15B). Because Cascade genes co-occur as conserved operons, they are typically arranged in a consistent order, within the same Type I subtype, facilitating their identification and selection for follow-up analysis and experimentation. As an example, additional Type I-E systems can be identified by locating Cas8 homologs, identifying promising bacterial species for homologous Cascade testing, and then obtaining or designing polynucleotide sequences encoding the Cas8 and other protein components of the Cascade from those homologous CRISPR-Cas systems.

[00175] Polynucleotide DNA sequences of genes encoding the subunit proteins of Cascade from a number of species (listed in Table 3 and Table 4), some with Cascade complexes homologous to those derived from *E. coli* strain K-12 MG1655, and the amino acid sequences of the resulting protein components, as well as exemplary minimal CRISPR arrays, are presented as SEQ ID NO:22 through SEQ ID NO:213 (Table 3).

[00176]

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 1	Cas8	I-E_Escherichia coli K-12 MG1655	Genomic DNA gene sequence
SEQ ID NO: 2	Cse2	I-E_Escherichia coli K-12 MG1655	Genomic DNA gene sequence
SEQ ID NO: 3	Cas7	I-E_Escherichia coli K-12 MG1655	Genomic DNA gene sequence
SEQ ID NO: 4	Cas5	I-E_Escherichia coli K-12 MG1655	Genomic DNA gene sequence
SEQ ID NO: 5	Cas6	I-E_Escherichia coli K-12 MG1655	Genomic DNA gene sequence
SEQ ID NO: 6	Cas8	I-E_Escherichia coli K-12 MG1655	*E. coli* codon-optimized DNA gene sequence
SEQ ID NO: 7	Cse2	I-E_Escherichia coli K-12 MG1655	*E. coli* codon-optimized DNA gene sequence
SEQ ID NO: 8	Cas7	I-E_Escherichia coli K-12 MG1655	*E. coli* codon-optimized DNA gene sequence
SEQ ID NO: 9	Cas5	I-E_Escherichia coli K-12 MG1655	*E. coli* codon-optimized DNA gene sequence
SEQ ID NO: 10	Cas6	I-E_Escherichia coli K-12 MG1655	*E. coli* codon-optimized DNA gene sequence
SEQ ID NO: 11	Cas8	I-E_Escherichia coli K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 12	Cse2	I-E_Escherichia coli K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 13	Cas7	I-E_Escherichia coli K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 14	Cas5	I-E_Escherichia coli K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 15	Cas6	I-E_Escherichia coli K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 16	Cas8	I-E_Escherichia coli K-12 MG1655	Protein amino acid sequence
SEQ ID NO: 17	Cse2	I-E_Escherichia coli K-12 MG1655	Protein amino acid sequence
SEQ ID NO: 18	Cas7	I-E_Escherichia coli K-12 MG1655	Protein amino acid sequence
SEQ ID NO: 19	Cas5	I-E_Escherichia coli K-12 MG1655	Protein amino acid sequence
SEQ ID NO: 20	Cas6	I-E_Escherichia coli K-12 MG1655	Protein amino acid sequence
SEQ ID NO: 21	Cas3	I-E_Escherichia coli K-12 MG1655	Protein amino acid sequence
SEQ ID NO: 22	Cas8	I-E_Oceanicola sp. HL-35	Genomic DNA gene sequence
SEQ ID NO: 23	Cse2	I-E_Oceanicola sp. HL-35	Genomic DNA gene sequence
SEQ ID NO: 24	Cas7	I-E_Oceanicola sp. HL-35	Genomic DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 25	Cas5	I-E_Oceanicola sp. HL-35	Genomic DNA gene sequence
SEQ ID NO: 26	Cas6	I-E_Oceanicola sp. HL-35	Genomic DNA gene sequence
SEQ ID NO: 27	Cas8	I-E_Oceanicola sp. HL-35	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 28	Cse2	I-E_Oceanicola sp. HL-35	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 29	Cas7	I-E_Oceanicola sp. HL-35	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 30	Cas5	I-E_Oceanicola sp. HL-35	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 31	Cas6	I-E_Oceanicola sp. HL-35	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 32	Cas8	I-E_Oceanicola sp. HL-35	Protein amino acid sequence
SEQ ID NO: 33	Cse2	I-E_Oceanicola sp. HL-35	Protein amino acid sequence
SEQ ID NO: 34	Cas7	I-E_Oceanicola sp. HL-35	Protein amino acid sequence
SEQ ID NO: 35	Cas5	I-E_Oceanicola sp. HL-35	Protein amino acid sequence
SEQ ID NO: 36	Cas6	I-E_Oceanicola sp. HL-35	Protein amino acid sequence
SEQ ID NO: 37	CRISPR	I-E_Oceanicola sp. HL-35	Exemplary minimal CRISPR array
SEQ ID NO: 38	Cas8	I-E_Pseudomonas sp. S-6-2	Genomic DNA gene sequence
SEQ ID NO: 39	Cse2	I-E_Pseudomonas sp. S-6-2	Genomic DNA gene sequence
SEQ ID NO: 40	Cas7	I-E_Pseudomonas sp. S-6-2	Genomic DNA gene sequence
SEQ ID NO: 41	Cas5	I-E_Pseudomonas sp. S-6-2	Genomic DNA gene sequence
SEQ ID NO: 42	Cas6	I-E_Pseudomonas sp. S-6-2	Genomic DNA gene sequence
SEQ ID NO: 43	Cas8	I-E_Pseudomonas sp. S-6-2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 44	Cse2	I-E_Pseudomonas sp. S-6-2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 45	Cas7	I-E_Pseudomonas sp. S-6-2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 46	Cas5	I-E_Pseudomonas sp. S-6-2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 47	Cas6	I-E_Pseudomonas sp. S-6-2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 48	Cas8	I-E_Pseudomonas sp. S-6-2	Protein amino acid sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 49	Cse2	I-E_Pseudomonas sp. S-6-2	Protein amino acid sequence
SEQ ID NO: 50	Cas7	I-E_Pseudomonas sp. S-6-2	Protein amino acid sequence
SEQ ID NO: 51	Cas5	I-E_Pseudomonas sp. S-6-2	Protein amino acid sequence
SEQ ID NO: 52	Cas6	I-E_Pseudomonas sp. S-6-2	Protein amino acid sequence
SEQ ID NO: 53	CRISPR	I-E_Pseudomonas sp. S-6-2	Exemplary minimal CRISPR array
SEQ ID NO: 54	Cas8	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Genomic DNA gene sequence
SEQ ID NO: 55	Cse2	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Genomic DNA gene sequence
SEQ ID NO: 56	Cas7	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Genomic DNA gene sequence
SEQ ID NO: 57	Cas5	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Genomic DNA gene sequence
SEQ ID NO: 58	Cas6	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Genomic DNA gene sequence
SEQ ID NO: 59	Cas8	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 60	Cse2	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 61	Cas7	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 62	Cas5	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 63	Cas6	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 64	Cas8	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Protein amino acid sequence
SEQ ID NO: 65	Cse2	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Protein amino acid sequence
SEQ ID NO: 66	Cas7	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Protein amino acid sequence
SEQ ID NO: 67	Cas5	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Protein amino acid sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 68	Cas6	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Protein amino acid sequence
SEQ ID NO: 69	CRISPR	I-E_Salmonella enterica subsp. enterica serovar Muenster strain	Exemplary minimal CRISPR array
SEQ ID NO: 70	Cas8	I-E_Atlantibacter hermannii NBRC 105704	Genomic DNA gene sequence
SEQ ID NO: 71	Cse2	I-E_Atlantibacter hermannii NBRC 105704	Genomic DNA gene sequence
SEQ ID NO: 72	Cas7	I-E_Atlantibacter hermannii NBRC 105704	Genomic DNA gene sequence
SEQ ID NO: 73	Cas5	I-E_Atlantibacter hermannii NBRC 105704	Genomic DNA gene sequence
SEQ ID NO: 74	Cas6	I-E_Atlantibacter hermannii NBRC 105704	Genomic DNA gene sequence
SEQ ID NO: 75	Cas8	I-E_Atlantibacter hermannii NBRC 105704	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 76	Cse2	I-E_Atlantibacter hermannii NBRC 105704	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 77	Cas7	I-E_Atlantibacter hermannii NBRC 105704	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 78	Cas5	I-E_Atlantibacter hermannii NBRC 105704	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 79	Cas6	I-E_Atlantibacter hermannii NBRC 105704	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 80	Cas8	I-E_Atlantibacter hermannii NBRC 105704	Protein amino acid sequence
SEQ ID NO: 81	Cse2	I-E_Atlantibacter hermannii NBRC 105704	Protein amino acid sequence
SEQ ID NO: 82	Cas7	I-E_Atlantibacter hermannii NBRC 105704	Protein amino acid sequence
SEQ ID NO: 83	Cas5	I-E_Atlantibacter hermannii NBRC 105704	Protein amino acid sequence
SEQ ID NO: 84	Cas6	I-E_Atlantibacter hermannii NBRC 105704	Protein amino acid sequence
SEQ ID NO: 85	CRISPR	I-E_Atlantibacter hermannii NBRC 105704	Exemplary minimal CRISPR array
SEQ ID NO: 86	Cas8	I-E_Geothermobacter sp. EPR-M	Genomic DNA gene sequence
SEQ ID NO: 87	Cse2	I-E_Geothermobacter sp. EPR-M	Genomic DNA gene sequence
SEQ ID NO: 88	Cas7	I-E_Geothermobacter sp. EPR-M	Genomic DNA gene sequence
SEQ ID NO: 89	Cas5	I-E_Geothermobacter sp. EPR-M	Genomic DNA gene sequence
SEQ ID NO: 90	Cas6	I-E_Geothermobacter sp. EPR-M	Genomic DNA gene sequence
SEQ ID NO: 91	Cas8	I-E_Geothermobacter sp. EPR-M	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 92	Cse2	I-E_Geothermobacter sp. EPR-M	*H. sapiens* codon-optimized DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 93	Cas7	I-E_Geothermobacter sp. EPR-M	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 94	Cas5	I-E_Geothermobacter sp. EPR-M	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 95	Cas6	I-E_Geothermobacter sp. EPR-M	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 96	Cas8	I-E_Geothermobacter sp. EPR-M	Protein amino acid sequence
SEQ ID NO: 97	Cse2	I-E_Geothermobacter sp. EPR-M	Protein amino acid sequence
SEQ ID NO: 98	Cas7	I-E_Geothermobacter sp. EPR-M	Protein amino acid sequence
SEQ ID NO: 99	Cas5	I-E_Geothermobacter sp. EPR-M	Protein amino acid sequence
SEQ ID NO: 100	Cas6	I-E_Geothermobacter sp. EPR-M	Protein amino acid sequence
SEQ ID NO: 101	CRISPR	I-E_Geothermobacter sp. EPR-M	Exemplary minimal CRISPR array
SEQ ID NO: 102	Cas8	I-E_Methylocaldum sp. 14B	Genomic DNA gene sequence
SEQ ID NO: 103	Cse2	I-E_Methylocaldum sp. 14B	Genomic DNA gene sequence
SEQ ID NO: 104	Cas7	I-E_Methylocaldum sp. 14B	Genomic DNA gene sequence
SEQ ID NO: 105	Cas5	I-E_Methylocaldum sp. 14B	Genomic DNA gene sequence
SEQ ID NO: 106	Cas6	I-E_Methylocaldum sp. 14B	Genomic DNA gene sequence
SEQ ID NO: 107	Cas8	I-E_Methylocaldum sp. 14B	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 108	Cse2	I-E_Methylocaldum sp. 14B	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 109	Cas7	I-E_Methylocaldum sp. 14B	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 110	Cas5	I-E_Methylocaldum sp. 14B	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 111	Cas6	I-E_Methylocaldum sp. 14B	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 112	Cas8	I-E_Methylocaldum sp. 14B	Protein amino acid sequence
SEQ ID NO: 113	Cse2	I-E_Methylocaldum sp. 14B	Protein amino acid sequence
SEQ ID NO: 114	Cas7	I-E_Methylocaldum sp. 14B	Protein amino acid sequence
SEQ ID NO: 115	Cas5	I-E_Methylocaldum sp. 14B	Protein amino acid sequence
SEQ ID NO: 116	Cas6	I-E_Methylocaldum sp. 14B	Protein amino acid sequence
SEQ ID NO: 117	CRISPR	I-E_Methylocaldum sp. 14B	Exemplary minimal CRISPR array
SEQ ID NO: 118	Cas8	I-E_Methanocella arvoryzae MRE50	Genomic DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 119	Cse2	I-E_Methanocella arvoryzae MRE50	Genomic DNA gene sequence
SEQ ID NO: 120	Cas7	I-E_Methanocella arvoryzae MRE50	Genomic DNA gene sequence
SEQ ID NO: 121	Cas5	I-E_Methanocella arvoryzae MRE50	Genomic DNA gene sequence
SEQ ID NO: 122	Cas6	I-E_Methanocella arvoryzae MRE50	Genomic DNA gene sequence
SEQ ID NO: 123	Cas8	I-E_Methanocella arvoryzae MRE50	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 124	Cse2	I-E_Methanocella arvoryzae MRE50	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 125	Cas7	I-E_Methanocella arvoryzae MRE50	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 126	Cas5	I-E_Methanocella arvoryzae MRE50	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 127	Cas6	I-E_Methanocella arvoryzae MRE50	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 128	Cas8	I-E_Methanocella arvoryzae MRE50	Protein amino acid sequence
SEQ ID NO: 129	Cse2	I-E_Methanocella arvoryzae MRE50	Protein amino acid sequence
SEQ ID NO: 130	Cas7	I-E_Methanocella arvoryzae MRE50	Protein amino acid sequence
SEQ ID NO: 131	Cas5	I-E_Methanocella arvoryzae MRE50	Protein amino acid sequence
SEQ ID NO: 132	Cas6	I-E_Methanocella arvoryzae MRE50	Protein amino acid sequence
SEQ ID NO: 133	CRISPR	I-E_Methanocella arvoryzae MRE50	Exemplary minimal CRISPR array
SEQ ID NO: 134	Cas8	I-E_Lachnospiraceae bacterium KH1T2	Genomic DNA gene sequence
SEQ ID NO: 135	Cse2	I-E_Lachnospiraceae bacterium KH1T2	Genomic DNA gene sequence
SEQ ID NO: 136	Cas7	I-E_Lachnospiraceae bacterium KH1T2	Genomic DNA gene sequence
SEQ ID NO: 137	Cas5	I-E_Lachnospiraceae bacterium KH1T2	Genomic DNA gene sequence
SEQ ID NO: 138	Cas6	I-E_Lachnospiraceae bacterium KH1T2	Genomic DNA gene sequence
SEQ ID NO: 139	Cas8	I-E_Lachnospiraceae bacterium KH1T2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 140	Cse2	I-E_Lachnospiraceae bacterium KH1T2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 141	Cas7	I-E_Lachnospiraceae bacterium KH1T2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 142	Cas5	I-E_Lachnospiraceae bacterium KH1T2	*H. sapiens* codon-optimized DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 143	Cas6	I-E_Lachnospiraceae bacterium KH1T2	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 144	Cas8	I-E_Lachnospiraceae bacterium KH1T2	Protein amino acid sequence
SEQ ID NO: 145	Cse2	I-E_Lachnospiraceae bacterium KH1T2	Protein amino acid sequence
SEQ ID NO: 146	Cas7	I-E_Lachnospiraceae bacterium KH1T2	Protein amino acid sequence
SEQ ID NO: 147	Cas5	I-E_Lachnospiraceae bacterium KH1T2	Protein amino acid sequence
SEQ ID NO: 148	Cas6	I-E_Lachnospiraceae bacterium KH1T2	Protein amino acid sequence
SEQ ID NO: 149	CRISPR	I-E_Lachnospiraceae bacterium KH1T2	Exemplary minimal CRISPR array
SEQ ID NO: 150	Cas8	I-E_Klebsiella pneumoniae strain VRCO0172	Genomic DNA gene sequence
SEQ ID NO: 151	Cse2	I-E_Klebsiella pneumoniae strain VRCO0172	Genomic DNA gene sequence
SEQ ID NO: 152	Cas7	I-E_Klebsiella pneumoniae strain VRCO0172	Genomic DNA gene sequence
SEQ ID NO: 153	Cas5	I-E_Klebsiella pneumoniae strain VRCO0172	Genomic DNA gene sequence
SEQ ID NO: 154	Cas6	I-E_Klebsiella pneumoniae strain VRCO0172	Genomic DNA gene sequence
SEQ ID NO: 155	Cas8	I-E_Klebsiella pneumoniae strain VRCO0172	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 156	Cse2	I-E_Klebsiella pneumoniae strain VRCO0172	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 157	Cas7	I-E_Klebsiella pneumoniae strain VRCO0172	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 158	Cas5	I-E_Klebsiella pneumoniae strain VRCO0172	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 159	Cas6	I-E_Klebsiella pneumoniae strain VRCO0172	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 160	Cas8	I-E_Klebsiella pneumoniae strain VRCO0172	Protein amino acid sequence
SEQ ID NO: 161	Cse2	I-E_Klebsiella pneumoniae strain VRCO0172	Protein amino acid sequence
SEQ ID NO: 162	Cas7	I-E_Klebsiella pneumoniae strain VRCO0172	Protein amino acid sequence
SEQ ID NO: 163	Cas5	I-E_Klebsiella pneumoniae strain VRCO0172	Protein amino acid sequence
SEQ ID NO: 164	Cas6	I-E_Klebsiella pneumoniae strain VRCO0172	Protein amino acid sequence
SEQ ID NO: 165	CRISPR	I-E_Klebsiella pneumoniae strain VRCO0172	Exemplary minimal CRISPR array
SEQ ID NO: 166	Cas8	I-E_Pseudomonas aeruginosa DHS01	Genomic DNA gene sequence
SEQ ID NO: 167	Cse2	I-E_Pseudomonas aeruginosa DHS01	Genomic DNA gene sequence
SEQ ID NO: 168	Cas7	I-E_Pseudomonas aeruginosa DHS01	Genomic DNA gene sequence
SEQ ID NO: 169	Cas5	I-E_Pseudomonas aeruginosa DHS01	Genomic DNA gene sequence
SEQ ID NO: 170	Cas6	I-E_Pseudomonas aeruginosa DHS01	Genomic DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 171	Cas8	I-E_Pseudomonas aeruginosa DHS01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 172	Cse2	I-E_Pseudomonas aeruginosa DHS01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 173	Cas7	I-E_Pseudomonas aeruginosa DHS01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 174	Cas5	I-E_Pseudomonas aeruginosa DHS01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 175	Cas6	I-E_Pseudomonas aeruginosa DHS01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 176	Cas8	I-E_Pseudomonas aeruginosa DHS01	Protein amino acid sequence
SEQ ID NO: 177	Cse2	I-E_Pseudomonas aeruginosa DHS01	Protein amino acid sequence
SEQ ID NO: 178	Cas7	I-E_Pseudomonas aeruginosa DHS01	Protein amino acid sequence
SEQ ID NO: 179	Cas5	I-E_Pseudomonas aeruginosa DHS01	Protein amino acid sequence
SEQ ID NO: 180	Cas6	I-E_Pseudomonas aeruginosa DHS01	Protein amino acid sequence
SEQ ID NO: 181	CRISPR	I-E_Pseudomonas aeruginosa DHS01	Exemplary minimal CRISPR array
SEQ ID NO: 182	Cas8	I-E_Streptococcus thermophilus strain ND07	Genomic DNA gene sequence
SEQ ID NO: 183	Cse2	I-E_Streptococcus thermophilus strain ND07	Genomic DNA gene sequence
SEQ ID NO: 184	Cas7	I-E_Streptococcus thermophilus strain ND07	Genomic DNA gene sequence
SEQ ID NO: 185	Cas5	I-E_Streptococcus thermophilus strain ND07	Genomic DNA gene sequence
SEQ ID NO: 186	Cas6	I-E_Streptococcus thermophilus strain ND07	Genomic DNA gene sequence
SEQ ID NO: 187	Cas8	I-E_Streptococcus thermophilus strain ND07	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 188	Cse2	I-E_Streptococcus thermophilus strain ND07	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 189	Cas7	I-E_Streptococcus thermophilus strain ND07	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 190	Cas5	I-E_Streptococcus thermophilus strain ND07	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 191	Cas6	I-E_Streptococcus thermophilus strain ND07	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 192	Cas8	I-E_Streptococcus thermophilus strain ND07	Protein amino acid sequence
SEQ ID NO: 193	Cse2	I-E_Streptococcus thermophilus strain ND07	Protein amino acid sequence
SEQ ID NO: 194	Cas7	I-E_Streptococcus thermophilus strain ND07	Protein amino acid sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 195	Cas5	I-E_Streptococcus thermophilus strain ND07	Protein amino acid sequence
SEQ ID NO: 196	Cas6	I-E_Streptococcus thermophilus strain ND07	Protein amino acid sequence
SEQ ID NO: 197	CRISPR	I-E_Streptococcus thermophilus strain ND07	Exemplary minimal CRISPR array
SEQ ID NO: 198	Cas8	I-E_Streptomyces sp. S4	Genomic DNA gene sequence
SEQ ID NO: 199	Cse2	I-E_Streptomyces sp. S4	Genomic DNA gene sequence
SEQ ID NO: 200	Cas7	I-E_Streptomyces sp. S4	Genomic DNA gene sequence
SEQ ID NO: 201	Cas5	I-E_Streptomyces sp. S4	Genomic DNA gene sequence
SEQ ID NO: 202	Cas6	I-E_Streptomyces sp. S4	Genomic DNA gene sequence
SEQ ID NO: 203	Cas8	I-E_Streptomyces sp. S4	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 204	Cse2	I-E_Streptomyces sp. S4	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 205	Cas7	I-E_Streptomyces sp. S4	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 206	Cas5	I-E_Streptomyces sp. S4	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 207	Cas6	I-E_Streptomyces sp. S4	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 208	Cas8	I-E_Streptomyces sp. S4	Protein amino acid sequence
SEQ ID NO: 209	Cse2	I-E_Streptomyces sp. S4	Protein amino acid sequence
SEQ ID NO: 210	Cas7	I-E_Streptomyces sp. S4	Protein amino acid sequence
SEQ ID NO: 211	Cas5	I-E_Streptomyces sp. S4	Protein amino acid sequence
SEQ ID NO: 212	Cas6	I-E_Streptomyces sp. S4	Protein amino acid sequence
SEQ ID NO: 213	CRISPR	I-E_Streptomyces sp. S4	Exemplary minimal CRISPR array
SEQ ID NO: 214	Cas6	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Genomic DNA gene sequence
SEQ ID NO: 215	Cas8	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Genomic DNA gene sequence
SEQ ID NO: 216	Cas7	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Genomic DNA gene sequence
SEQ ID NO: 217	Cas5	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Genomic DNA gene sequence
SEQ ID NO: 218	Cas6	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	*H. sapiens* codon-optimized DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 219	Cas8	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 220	Cas7	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 221	Cas5	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 222	Cas6	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Protein amino acid sequence
SEQ ID NO: 223	Cas8	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Protein amino acid sequence
SEQ ID NO: 224	Cas7	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Protein amino acid sequence
SEQ ID NO: 225	Cas5	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Protein amino acid sequence
SEQ ID NO: 226	CRISPR	I-B_Fusobacterium nucleatum subsp. animalis 3_1_33	Exemplary minimal CRISPR array
SEQ ID NO: 227	Cas6	I-B_Campylobacter fetus subsp. testudinum Sp3	Genomic DNA gene sequence
SEQ ID NO: 228	Cas8	I-B_Campylobacter fetus subsp. testudinum Sp3	Genomic DNA gene sequence
SEQ ID NO: 229	Cas7	I-B_Campylobacter fetus subsp. testudinum Sp3	Genomic DNA gene sequence
SEQ ID NO: 230	Cas5	I-B_Campylobacter fetus subsp. testudinum Sp3	Genomic DNA gene sequence
SEQ ID NO: 231	Cas6	I-B_Campylobacter fetus subsp. testudinum Sp3	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 232	Cas8	I-B_Campylobacter fetus subsp. testudinum Sp3	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 233	Cas7	I-B_Campylobacter fetus subsp. testudinum Sp3	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 234	Cas5	I-B_Campylobacter fetus subsp. testudinum Sp3	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 235	Cas6	I-B_Campylobacter fetus subsp. testudinum Sp3	Protein amino acid sequence
SEQ ID NO: 236	Cas8	I-B_Campylobacter fetus subsp. testudinum Sp3	Protein amino acid sequence
SEQ ID NO: 237	Cas7	I-B_Campylobacter fetus subsp. testudinum Sp3	Protein amino acid sequence
SEQ ID NO: 238	Cas5	I-B_Campylobacter fetus subsp. testudinum Sp3	Protein amino acid sequence
SEQ ID NO: 239	CRISPR	I-B_Campylobacter fetus subsp. testudinum Sp3	Exemplary minimal CRISPR array
SEQ ID NO: 240	Cas6	I-B_Odoribacter splanchnicus DSM 20712	Genomic DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 241	Cas8	I-B_Odoribacter splanchnicus DSM 20712	Genomic DNA gene sequence
SEQ ID NO: 242	Cas7	I-B_Odoribacter splanchnicus DSM 20712	Genomic DNA gene sequence
SEQ ID NO: 243	Cas5	I-B_Odoribacter splanchnicus DSM 20712	Genomic DNA gene sequence
SEQ ID NO: 244	Cas6	I-B_Odoribacter splanchnicus DSM 20712	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 245	Cas8	I-B_Odoribacter splanchnicus DSM 20712	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 246	Cas7	I-B_Odoribacter splanchnicus DSM 20712	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 247	Cas5	I-B_Odoribacter splanchnicus DSM 20712	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 248	Cas6	I-B_Odoribacter splanchnicus DSM 20712	Protein amino acid sequence
SEQ ID NO: 249	Cas8	I-B_Odoribacter splanchnicus DSM 20712	Protein amino acid sequence
SEQ ID NO: 250	Cas7	I-B_Odoribacter splanchnicus DSM 20712	Protein amino acid sequence
SEQ ID NO: 251	Cas5	I-B_Odoribacter splanchnicus DSM 20712	Protein amino acid sequence
SEQ ID NO: 252	CRISPR	I-B_Odoribacter splanchnicus DSM 20712	Exemplary minimal CRISPR array
SEQ ID NO: 253	Cas5	I-C_Bacillus halodurans C-125	Genomic DNA gene sequence
SEQ ID NO: 254	Cas8	I-C_Bacillus halodurans C-125	Genomic DNA gene sequence
SEQ ID NO: 255	Cas7	I-C_Bacillus halodurans C-125	Genomic DNA gene sequence
SEQ ID NO: 256	Cas5	I-C_Bacillus halodurans C-125	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 257	Cas8	I-C_Bacillus halodurans C-125	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 258	Cas7	I-C_Bacillus halodurans C-125	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 259	Cas5	I-C_Bacillus halodurans C-125	Protein amino acid sequence
SEQ ID NO: 260	Cas8	I-C_Bacillus halodurans C-125	Protein amino acid sequence
SEQ ID NO: 261	Cas7	I-C_Bacillus halodurans C-125	Protein amino acid sequence
SEQ ID NO: 262	CRISPR	I-C_Bacillus halodurans C-125	Exemplary minimal CRISPR array
SEQ ID NO: 263	Cas5	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	Genomic DNA gene sequence
SEQ ID NO: 264	Cas8	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	Genomic DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 265	Cas7	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	Genomic DNA gene sequence
SEQ ID NO: 266	Cas5	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 267	Cas8	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 268	Cas7	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 269	Cas5	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	Protein amino acid sequence
SEQ ID NO: 270	Cas8	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	Protein amino acid sequence
SEQ ID NO: 271	Cas7	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	Protein amino acid sequence
SEQ ID NO: 272	CRISPR	I-C_Desulfovibrio vulgaris RCH1 plasmid pDEVAL01	Exemplary minimal CRISPR array
SEQ ID NO: 273	Cas5	I-C_Geobacillus thermocatenulatus strain KCTC 3921	Genomic DNA gene sequence
SEQ ID NO: 274	Cas8	I-C_Geobacillus thermocatenulatus strain KCTC 3921	Genomic DNA gene sequence
SEQ ID NO: 275	Cas7	I-C_Geobacillus thermocatenulatus strain KCTC 3921	Genomic DNA gene sequence
SEQ ID NO: 276	Cas5	I-C_Geobacillus thermocatenulatus strain KCTC 3921	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 277	Cas8	I-C_Geobacillus thermocatenulatus strain KCTC 3921	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 278	Cas7	I-C_Geobacillus thermocatenulatus strain KCTC 3921	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 279	Cas5	I-C_Geobacillus thermocatenulatus strain KCTC 3921	Protein amino acid sequence
SEQ ID NO: 280	Cas8	I-C_Geobacillus thermocatenulatus strain KCTC 3921	Protein amino acid sequence
SEQ ID NO: 281	Cas7	I-C_Geobacillus thermocatenulatus strain KCTC 3921	Protein amino acid sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 282	CRISPR	I-C_Geobacillus thermocatenulatus strain KCTC 3921	Exemplary minimal CRISPR array
SEQ ID NO: 283	Cas8	I-F_Vibrio cholerae strain L15	Genomic DNA gene sequence
SEQ ID NO: 284	Cas5	I-F_Vibrio cholerae strain L15	Genomic DNA gene sequence
SEQ ID NO: 285	Cas7	I-F_Vibrio cholerae strain L15	Genomic DNA gene sequence
SEQ ID NO: 286	Cas6	I-F_Vibrio cholerae strain L15	Genomic DNA gene sequence
SEQ ID NO: 287	Cas8	I-F_Vibrio cholerae strain L15	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 288	Cas5	I-F_Vibrio cholerae strain L15	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 289	Cas7	I-F_Vibrio cholerae strain L15	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 290	Cas6	I-F_Vibrio cholerae strain L15	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 291	Cas8	I-F_Vibrio cholerae strain L15	Protein amino acid sequence
SEQ ID NO: 292	Cas5	I-F_Vibrio cholerae strain L15	Protein amino acid sequence
SEQ ID NO: 293	Cas7	I-F_Vibrio cholerae strain L15	Protein amino acid sequence
SEQ ID NO: 294	Cas6	I-F_Vibrio cholerae strain L15	Protein amino acid sequence
SEQ ID NO: 295	CRISPR	I-F_Vibrio cholerae strain L15	Exemplary minimal CRISPR array
SEQ ID NO: 296	Cas8	I-F_Klebsiella oxytoca strain ICU1-2b	Genomic DNA gene sequence
SEQ ID NO: 297	Cas5	I-F_Klebsiella oxytoca strain ICU1-2b	Genomic DNA gene sequence
SEQ ID NO: 298	Cas7	I-F_Klebsiella oxytoca strain ICU1-2b	Genomic DNA gene sequence
SEQ ID NO: 299	Cas6	I-F_Klebsiella oxytoca strain ICU1-2b	Genomic DNA gene sequence
SEQ ID NO: 300	Cas8	I-F_Klebsiella oxytoca strain ICU1-2b	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 301	Cas5	I-F_Klebsiella oxytoca strain ICU1-2b	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 302	Cas7	I-F_Klebsiella oxytoca strain ICU1-2b	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 303	Cas6	I-F_Klebsiella oxytoca strain ICU1-2b	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 304	Cas8	I-F_Klebsiella oxytoca strain ICU1-2b	Protein amino acid sequence
SEQ ID NO: 305	Cas5	I-F_Klebsiella oxytoca strain ICU1-2b	Protein amino acid sequence
SEQ ID NO: 306	Cas7	I-F_Klebsiella oxytoca strain ICU1-2b	Protein amino acid sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 307	Cas6	I-F_Klebsiella oxytoca strain ICU1-2b	Protein amino acid sequence
SEQ ID NO: 308	CRISPR	I-F_Klebsiella oxytoca strain ICU1-2b	Exemplary minimal CRISPR array
SEQ ID NO: 309	Cas8	I-F_Pseudomonas aeruginosa UCBPP-PA14	Genomic DNA gene sequence
SEQ ID NO: 310	Cas5	I-F_Pseudomonas aeruginosa UCBPP-PA14	Genomic DNA gene sequence
SEQ ID NO: 311	Cas7	I-F_Pseudomonas aeruginosa UCBPP-PA14	Genomic DNA gene sequence
SEQ ID NO: 312	Cas6	I-F_Pseudomonas aeruginosa UCBPP-PA14	Genomic DNA gene sequence
SEQ ID NO: 313	Cas8	I-F_Pseudomonas aeruginosa UCBPP-PA14	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 314	Cas5	I-F_Pseudomonas aeruginosa UCBPP-PA14	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 315	Cas7	I-F_Pseudomonas aeruginosa UCBPP-PA14	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 316	Cas6	I-F_Pseudomonas aeruginosa UCBPP-PA14	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 317	Cas8	I-F_Pseudomonas aeruginosa UCBPP-PA14	Protein amino acid sequence
SEQ ID NO: 318	Cas5	I-F_Pseudomonas aeruginosa UCBPP-PA14	Protein amino acid sequence
SEQ ID NO: 319	Cas7	I-F_Pseudomonas aeruginosa UCBPP-PA14	Protein amino acid sequence
SEQ ID NO: 320	Cas6	I-F_Pseudomonas aeruginosa UCBPP-PA14	Protein amino acid sequence
SEQ ID NO: 321	CRISPR	I-F_Pseudomonas aeruginosa UCBPP-PA14	Exemplary minimal CRISPR array
SEQ ID NO: 322	Cas7	I-Fv2_Shewanella putrefaciens CN-32	Genomic DNA gene sequence
SEQ ID NO: 323	Cas5	I-Fv2_Shewanella putrefaciens CN-32	Genomic DNA gene sequence
SEQ ID NO: 324	Cas6	I-Fv2_Shewanella putrefaciens CN-32	Genomic DNA gene sequence
SEQ ID NO: 325	Cas7	I-Fv2_Shewanella putrefaciens CN-32	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 326	Cas5	I-Fv2_Shewanella putrefaciens CN-32	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 327	Cas6	I-Fv2_Shewanella putrefaciens CN-32	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 328	Cas7	I-Fv2_Shewanella putrefaciens CN-32	Protein amino acid sequence
SEQ ID NO: 329	Cas5	I-Fv2_Shewanella putrefaciens CN-32	Protein amino acid sequence
SEQ ID NO: 330	Cas6	I-Fv2_Shewanella putrefaciens CN-32	Protein amino acid sequence
SEQ ID NO: 331	CRISPR	I-Fv2_Shewanella putrefaciens CN-32	Exemplary minimal CRISPR array
SEQ ID NO: 332	Cas7	I-Fv2_Acinetobacter sp. 869535	Genomic DNA gene sequence
SEQ ID NO: 333	Cas5	I-Fv2_Acinetobacter sp. 869535	Genomic DNA gene sequence

Table 3
Polynucleotide Sequences for Genes Encoding Cascade Proteins From 12 Species

SEQ ID NO:	Gene/ Protein	Subtype and organism	Type of sequence
SEQ ID NO: 334	Cas6	I-Fv2_Acinetobacter sp. 869535	Genomic DNA gene sequence
SEQ ID NO: 335	Cas7	I-Fv2_Acinetobacter sp. 869535	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 336	Cas5	I-Fv2_Acinetobacter sp. 869535	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 337	Cas6	I-Fv2_Acinetobacter sp. 869535	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 338	Cas7	I-Fv2_Acinetobacter sp. 869535	Protein amino acid sequence
SEQ ID NO: 339	Cas5	I-Fv2_Acinetobacter sp. 869535	Protein amino acid sequence
SEQ ID NO: 340	Cas6	I-Fv2_Acinetobacter sp. 869535	Protein amino acid sequence
SEQ ID NO: 341	CRISPR	I-Fv2_Acinetobacter sp. 869535	Exemplary minimal CRISPR array
SEQ ID NO: 342	Cas7	I-Fv2_Vibrio cholerae HE48	Genomic DNA gene sequence
SEQ ID NO: 343	Cas5	I-Fv2_Vibrio cholerae HE48	Genomic DNA gene sequence
SEQ ID NO: 344	Cas6	I-Fv2_Vibrio cholerae HE48	Genomic DNA gene sequence
SEQ ID NO: 345	Cas7	I-Fv2_Vibrio cholerae HE48	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 346	Cas5	I-Fv2_Vibrio cholerae HE48	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 347	Cas6	I-Fv2_Vibrio cholerae HE48	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO: 348	Cas7	I-Fv2_Vibrio cholerae HE48	Protein amino acid sequence
SEQ ID NO: 349	Cas5	I-Fv2_Vibrio cholerae HE48	Protein amino acid sequence
SEQ ID NO: 350	Cas6	I-Fv2_Vibrio cholerae HE48	Protein amino acid sequence
SEQ ID NO: 351	CRISPR	I-Fv2_Vibrio cholerae HE48	Exemplary minimal CRISPR array

[00177] The polynucleotide sequences for the proteins were derived from the gDNA of the host bacterium, and were codon-optimized specifically for expression in *E. coli*, and/or codon-optimized specifically for expression in eukaryotic cells (e.g., human cells). The polynucleotide DNA sequences encoding corresponding minimal CRISPR arrays were based on repeat sequences derived from the 12 species and can be used to generate mature crRNA that function as guide RNAs. In Table 4, the minimal CRISPR array comprises two repeat sequences (lower case, underlined) flanking an exemplary "spacer" sequence, which represents the guide portion of the crRNA. RNA processing by the endonuclease Cascade subunit generates a crRNA with repeat sequences on both the 5' and 3' ends, flanking the guide sequence.

[00178]

Table 4
Minimal CRISPR Arrays

SEQ ID NO:	Species	Minimal CRISPR repeat
SEQ ID NO:37	I-E_*Oceanicola sp.* HL-35	ctgttccccgcacacgcggggatgaaccgGGTTCT TCGATCTGCGCATCCATGATGCCGC Cctgttccccgcacacgcggggatgaaccg
SEQ ID NO:53	I-E_*Pseudomonas sp.* S-6-2	gtgttccccgcacctgcggggatgaaccGGGCCG GGGCGTTTGCGCTGTCAGGGGCGT CCCgtgttccccgcacctgcggggatgaaccg
SEQ ID NO:69	I-E_*Salmonella enterica subsp. enterica* serovar Muenster strain	gtgttccccgcgccagcggggataaaccgCAGCTT TAGCATCGGTCGACAGCCCATCTG GCgtgttccccgcgccagcggggataaaccg
SEQ ID NO:85	I-E_*Atlantibacter hermannii* NBRC 105704	gtgttccccgcgccagcggggataaaccgTTTTAA AACAGGATGTGGCCCGCCTGGTGC TGgtgttccccgcgccagcggggataaaccg
SEQ ID NO:101	I-E_*Geothermobacter sp.* EPR-M	ctgttccccgcacccgcggggatgaaccgGTCATC TATTTTTAATGGACGATATTTTTCA Actgttccccgcacccgcggggatgaaccg
SEQ ID NO:117	I-E_*Methylocaldum sp.* 14B	ctgttccccacgtacgtggggatgaaccgACGGCG TAATGGTAATTGTTAGCCGACAAG TTctgttccccacgtacgtggggatgaaccg
SEQ ID NO:133	I-E_*Methanocella arvoryzae* MRE50	aaagtccccacaggcgtgggggtgaaccgTGATC AGTAACCCGGTCACCATTAAACAG ATTaaagtccccacaggcgtgggggtgaaccg
SEQ ID NO:149	I-E_*Lachnospiraceae bacterium* KH1T2	gtattccccacgcacgtggrggtaaatcCGCTGAG TTTAATTACGCAGCGGAAGCCGGA GCGgtattccccacgcacgtgggggtaaatc
SEQ ID NO:165	I-E_*Klebsiella pneumoniae* strain VRCO0172	gtcttccccacacgcgtgggggtgtttcCGGCTCTT TTTTATCTCCTTCATCCTTCGCTATgt cttccccacacgcgtgggggtgtttc

Table 4 Minimal CRISPR Arrays		
SEQ ID NO:	**Species**	**Minimal CRISPR repeat**
SEQ ID NO:181	I-E_*Pseudomonas aeruginosa* DHS01	gtgttccccacatgcgtggggatgaaccgGGCACC ATCGGCGCCATTGACCGCGCGCTG AAGgtgttccccacatgcgtggggatgaaccg
SEQ ID NO:197	I-E_*Streptococcus thermophilus* strain ND07	gtttttcccgcacacgcgggggtgatccTATACCT ATATCAATGGCCTCCCACGCATAA GCgtttttcccgcacacgcgggggtgatcc
SEQ ID NO:213	I-E_*Streptomyces sp.* S4	gtcggccccgcacccgcggggatgctccAATGGC CGAGGACGACGGCGATCTGGCCAC GGACgtcggccccgcacccgcggggatgctcc

[00179] In another embodiment, the present invention relates to engineered polynucleotide sequences encoding Cascade components from additional bacterial or archaeal species, within other Type I subtypes; including, but not limited, to Types I-B, I-C, I-F, and variants of I-F, which can be identified and designed following the guidance of the present Specification and by using bioinformatics tools such as BLAST and PSI-BLAST to locate homologs of Cascade genes from hallmark systems typifying each subtype (see, e.g., Makarova, K.S., et al., Nat. Rev. Microbiol. 13:722-736 (2015); Koonin, E.V., et al., Curr. Opin. Microbiol. 37:67-78 (2017)). After identifying desirable homologs, the flanking genomic neighborhoods of the Cascade gene can be inspected to locate and identify genes of the remaining Cascade subunit proteins as disclosed herein. As an example, additional Type I-F systems can be identified by locating Cas8 homologs (and additional Type I-F variant 2 systems can be identified by locating Cas5 homologs) and identifying promising bacterial species for homologous Cascade testing, and then obtaining or designing polynucleotide sequences encoding the Cas8, Cas5, and other protein components of the Cascade from those homologous CRISPR-Cas systems.

[00180] Polynucleotide DNA sequences of genes encoding the three, four, or five subunit proteins of Cascade from Types I-B, I-C, I-F, and I-F variant 2 from 12 additional homologous Cascade complexes, and the amino acid sequences of the resulting protein components, as well as exemplary minimal CRISPR arrays, are presented as SEQ ID NO:214 through SEQ ID NO:351 (Table 3). The polynucleotide sequences for the subunit proteins were derived from the gDNA of the host bacterium, and were codon-optimized specifically for expression in *E. coli*, and/or codon-optimized specifically for expression in eukaryotic cells (e.g., human cells). The polynucleotide DNA sequences encoding corresponding minimal CRISPR arrays were based on repeat sequences derived from the 12 species and can be used to

generate mature crRNA that function as guide RNAs. In Table 5, the minimal CRISPR array comprises two repeat sequences (lower case, underlined) flanking an exemplary "spacer" sequence, which represents the guide portion of the crRNA. RNA processing by the endonuclease Cascade subunit generates a crRNA with repeat sequences on both the 5' and 3' ends, flanking the guide sequence.

[00181]

Table 5
Minimal CRISPR Arrays

SEQ ID NO:	Species	Minimal CRISPR repeat
SEQ ID NO:226	I-B_*Fusobacterium nucleatum subsp. animalis* 3_1_33	atgaactgtaaacttgaaaagttttgaaatGTTGACAA ATATTCAGATAATTTTTCAAAATCTT TTatgaactgtaaacttgaaaagttttgaaat
SEQ ID NO:239	I-B_*Campylobacter fetus subsp. testudinum* Sp3	gtttgctaatgacaatatttgtgttaaaacAAGCGTAG CACCAAAAGAAGCGTATGAAAGCAT AGgtttgctaatgacaatatttgtgttaaaac
SEQ ID NO:252	I-B_*Odoribacter splanchnicus* DSM 20712	cttttaattgaactaaggtagaattgaaacTAGGAATA AACCGTACCCAACCACGTAGCCATA TACGcttttaattgaactaaggtagaattgaaac
SEQ ID NO:262	I-C_*Bacillus halodurans* C-125	gtcgcactcttcatgggtgcgtggattgaaatCCTTTG ACGGAGAGGGGAACAGGAAATTAG AGAAGgtcgcactcttcatgggtgcgtggattgaaat
SEQ ID NO:272	I-C_*Desulfovibrio vulgaris* RCH1 plasmid pDEVAL01	gtcgccccccacgcgggggcgtggattgaaacCAGTC TCGTTACCCTGTCGCGGAGGGCGTCG ATgtcgccccccacgcgggggcgtggattgaaac
SEQ ID NO:282	I-C_*Geobacillus thermocatenulatus* strain KCTC 3921	gttgcacccggctattaagccgggtgaggattgaaacTA TATCACACAGCTTCTTAGTATCATCG ACAACACGTgttgcacccggctattaagccgggtg aggattgaaac
SEQ ID NO:295	I-F_*Vibrio cholerae* strain L15	gttcactgccgtacaggcagcttagaaaAATATGCA GGGGTTTGAAACGCTCGATGTTATgtt cactgccgtacaggcagcttagaaa
SEQ ID NO:308	I-F_*Klebsiella oxytoca* strain ICU1-2b	gttcactgccgtacaggcagcttagaaaAAAAACTG AGCGGCCGCAGAATGAAGTTGTAAgt tcactgccgtacaggcagcttagaaa
SEQ ID NO:321	I-F_*Pseudomonas aeruginosa* UCBPP-PA14	gttcactgccgtgtaggcagctaagaaaACCACCCG CTACCACCGGCAGCCGCACCGGCCgtt cactgccgtgtaggcagctaagaaa
SEQ ID NO:331	I-Fv2_*Shewanella putrefaciens* CN-32	gttcaccgccgcacaggcggcttagaaaTCAACCA AATCATAAATTGCGCGACCACATTGg ttcaccgccgcacaggcggcttagaaa
SEQ ID NO:341	I-Fv2_*Acinetobacter sp.* 869535	gttcactgccatataggcagcttagaaaATCGTTTTT TCATACGAGATTCGAAACGGACAgttc actgccatataggcagcttagaaa
SEQ ID NO:351	I-Fv2_*Vibrio cholerae* HE48	gttcactgccgcacaggcagcttagaaaTAACCGGA GGCGTACACTCGATAGAGGCAGCGgt tcactgccgcacaggcagcttagaaa

[00182] Example 19A to Example 19I and Example 22A to Example 22C describe the design and testing of multiple Cascade complex homologs, each comprising a Cas subunit protein-FokI fusion protein, to evaluate the efficiency of genome editing for each Cascade complex. The highest editing was observed with the variant from *Pseudomonas sp. S-6-2*, while other homologs (i.e., *Salmonella enterica*, *Geothermobacter sp.* EPR-M, *Methanocella arvoryzae MRE50*, and *S. thermophilus* (strain ND07)) showed editing approximately equivalent to *E. coli*. Editing was also observed with engineered *Vibrio cholera* strain L15 (Type I-F) FokI-Cascade complexes and *Vibrio cholera* strain HE48 (Type I-Fv2) FokI-Cascade complexes. In one embodiment, the different PAM requirements of these different homologs can increase target density in a target polynucleotide (e.g., gDNA in a cell). Accordingly, this collection of Cascade complex homologs provides greater flexibility in selection of nucleic acid target sequences in a target polynucleotide (e.g., gDNA in a cell).

[00183] In a second aspect, the present invention relates to modified Cascade subunit proteins. Cascade subunit proteins suitable for modification include, but are not limited to, Cascade subunit proteins of the species described herein.

[00184] In one embodiment, the present invention relates to engineered circular permutations of Cascade subunit proteins. Such circular permutations of a Cascade subunit protein result in a protein structure having different connectivity of the original linear sequence of amino acids of the Cascade subunit protein, but having an overall similar three-dimensional shape (see, e.g., Bliven, S., et al., PLoS Comput. Biol. 8:e1002445 (2012)). Circular permutations of Cascade subunit proteins can have a number of advantages. For example, a circular permutation of a Cas7 subunit protein can create a new N-terminus and a new C-terminus designed to be positioned for connection with an additional polypeptide sequence to form a fusion protein or linker region without disturbing the Cas7 protein fold or the Cascade complex assembly. Three examples of circular permutations of Cas7 (circularly permuted Cas7, cpCas7) are illustrated in FIG. 3A and FIG. 3B. In FIG. 3A and FIG. 3B, three portions of the protein are shown: an N-terminal portion of the native protein (FIG. 3A, vertical stripes, e.g., a Cas7 protein), a central portion of the native protein (FIG. 3A, grey shading), and a C-terminal portion of the native protein (FIG. 3A, no shading). FIG. 3A illustrates relocation of an N-terminal portion of the native protein to the C-terminal position of the native protein to produce a circularly permuted protein (FIG. 3A, cpCas7), wherein the N-terminal portion of the native protein is now at the N-terminal end of the cpCas7 and is connected to the central portion of the native protein by a linker polypeptide (FIG. 3A, Linker). FIG. 3B illustrates relocation of a C-

terminal portion of the native protein (FIG. 3B, Cas7) to the N-terminal position of the native protien (FIG. 3B, cpCas7), wherein the C-terminal portion of the native protein is now at the N-terminal end of the cpCas7 and is connected to the central portion of the native protein by a linker polypeptide (FIG. 3B, Linker).

[00185] The data presented in Example 10A, Example 10B, and Example 10 show that purification of Cascade complexes comprising circularly-permuted Cas7 subunit protein variants demonstrate that circularly-permuted Type I-E CRISPR-Cas subunit proteins can be successfully used to form Cascade complexes having essentially the same composition (based on molecular weight) as Cascade complexes comprising wild-type proteins.

[00186] In another embodiment, the present invention relates to Cascade subunit proteins fused to additional polypeptide sequences to create fusion proteins, as well as polynucleotides encoding such fusion proteins. Additional polypeptide sequences can include, but are not limited to, proteins, protein domains, protein fragments, and functional domains. Examples of such additional polypeptide sequences include, but are not limited to, sequences derived from transcription activator or repressor domains, and nucleotide deaminases (e.g., a cytidine deaminase or an adenine deaminase such as described in Komor, et. al., Nature 553:420-424 (2016); Koblan, et. al., Nat. Biotechnol. doi: 10.1038/nbt.4172 (May 29, 2018)). Additional functional domains for fusion proteins are presented herein.

[00187] An additional polypeptide sequence can be fused to any of the Cascade subunit proteins wherein the additional polypeptide sequence is encoded by an additional polynucleotide sequence that is typically appended to either the 5' or 3' end of a polynucleotide comprising the coding sequence of a Cascade subunit protein. In some embodiments, additional polynucleotide sequences that encode amino acid linkers connect a Cascade subunit protein to the additional polypeptide sequences of interest. In some embodiments, the polynucleotide sequences for the fusion protein partner and the linker sequence can be derived from naturally occurring gDNA sequences or may be codon-optimized for bacterial expression in *E. coli* or eukaryotic expression in mammalian cells (e.g., human cells). Examples of fusions proteins comprising affinity tags (e.g., His6, Strep-tag® II (IBA GMBH LLC, Göttingen, Germany)), nuclear localization signal or sequence (NLS), maltose binding protein, and FokI are presented in Example 1. Exemplary amino acid linker sequences are also disclosed in Example 1.

[00188] Example 11A describes Cascade subunit protein-FokI fusions, as well as Cascade subunit protein fusions to cytidine deaminases, endonucleases, restriction enzymes, nucleases/helicases, or domains thereof. Example 11B describes Cascade subunit protein fusions with other Cascade subunit proteins, as well as Cascade subunit protein fusions with

other Cascade subunit fusion proteins and an enzymatic protein domain (Example 11D). In some embodiments, a Type I CRISPR subunit protein can be evaluated in silico for the ability to be used to generate protein fusions at the N-terminus, C-terminus, or positions between the N-terminus and the C-terminus. In some embodiments, a Type I CRISPR subunit protein can be linked to one or more fusion domains at the N- terminus, C-terminus, or positions between the N-terminus and the C-terminus using one or more polypeptide linkers. In some embodiments, a Cascade subunit protein can be fused to a single-chain FokI (e.g., a single chain FokI fusion to EcoCascade RNP complex; nucleotide sequence, SEQ ID NO:1926; protein sequence, SEQ ID NO:1927). Exemplary polypeptide linkers are set forth in Examples 1, 11, 18, and 19.

[00189] FIG. 4A and FIG. 4B illustrate Cascade complexes comprising a Cas8 subunit protein (FIG. 4A, FIG. 4B, Cas7, Cas5, Cas8, Cse2, Cas6, the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin; and Cas8, "C" C-terminal, "N" N-terminal are indicated) fused to an additional protein sequence (e.g., a FokI). FIG. 4A shows an example of the additional protein sequence (FIG. 4A, FP) connected with the C-terminus of a Cas8 subunit protein using a linker polypeptide (FIG. 4A, black curved line). FIG. 4B shows an example of the additional protein sequence (FIG. 4B, FP) connected with the N-terminus of a Cas8 subunit protein using a linker polypeptide (FIG. 4B, black curved line). Example 11A describes in silico design, cloning, expression, and purification of a Type I-E Cas8 fused N-terminally with a FokI nuclease domain.

[00190] FIG. 5A and FIG. 5B illustrate additional examples of Cascade complexes comprising a Cascade subunit protein fused to an additional protein sequence. In FIG. 5A and FIG. 5B, the cRNA is illustrated as a black line comprising the hairpin and the relative positions of the Cas proteins of the Cascade complex are shown (FIG. 5A, FIG. 5B: Cas7, Cas5, Cas8, Cse2, Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin). FIG. 5A shows an example of a detectable moiety (e.g., a green fluorescent protein; FIG. 5A, GFP) fused to each of six Cas7 subunit proteins, each via a linker polypeptide (FIG. 5A, curved black line). Such a Cascade complex can be useful for detection of binding of the complex to a nucleic acid target sequence by providing significant signal amplification as a result of the presence of the multiple detectable moieties associated with the Cascade complex. FIG. 5B shows an example of an additional protein sequence (FIG. 5B, FP) connected with Cas6 subunit protein using a linker polypeptide (FIG. 5B, curved black line).

[00191] Examples of fusion proteins containing *E. coli* Type I-E Cascade subunit proteins include, but are not limited to, the following: the same subunit (e.g.,

Cse2_linker_Cse2), circularly permuted subunits (e.g., cpCas7_linker_cpCas7_linker_ cpCas7_linker_cpCas7_linker_cpCas7_linker_cpCas7), a Type I-E Cascade protein fused to a nuclease (e.g., FokI_linker_Cas8, Cas3_linker_Cas8, Cas6_linker_FokI, S1Nuclease_linker_ Cse2_linker_Cse2), a Type I-E Cascade protein fused to a cytidine deaminase (e.g., Cas8_linker_AID, Cse2_linker_Cse2_linker_APOBEC3G), and a Type I-E Cascade protein fused one or more other Type I-E Cascade proteins (e.g., Cas6_linker_cpCas7_linker_ cpCas7_ linker_cpCas7_linker_cpCas7_linker_cpCas7_ linker_cpCas7, cpCas7_linker_ cpCas7_linker_cpCas7_linker_cpCas7_linker_cpCas7_linker_cpCas7_linker_Cas5, Cas6_linker_cpCas7_linker_cpCas7_linker_cpCas7_linker_cpCas7_linker_cpCas7_linker_cpCas7_linker_cpCas7_linker_Cas5).

[00192] FIG. 6A, FIG. 6B, and FIG. 6C present illustrations of engineered Type I CRISPR-Cas effector complexes that contain cpCas7. In FIG. 6A, FIG. 6B, and FIG. 6C, "cpCas7" is a circularly permuted Cas7 protein (FIG. 6A, FIG. 6B, FIG. 6C: cpCas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin; for cpCas7 the shading corresponds to the circularly permuted protein illustrated in FIG. 3A), and the relative positions of the Cas proteins of the Cascade complex are shown. FIG. 6A presents a Cascade complex comprising six individual cpCas7 subunit proteins (FIG. 6A, cpCas7). FIG. 6B presents a Cascade complex comprising six fused cpCas7 subunit proteins, wherein the C-terminus of a cpCas7 subunit protein (FIG. 6B, cpCas7) is connected with the N-terminus of an adjacent cpCas7 subunit protein using a linker polypeptide (FIG. 6B, linker polypeptide is illustrated as a dark black line connecting the cpCas7 subunit proteins). FIG. 6C presents an embodiment wherein the Cascade complex comprises six fused cpCas7 subunit proteins (a "backbone"), wherein the C-terminus of the first cpCas7 subunit protein is connected with the N-terminus of the second cpCas7 subunit protein using a linker polypeptide (FIG. 6C, linker polypeptide is illustrated as a dark black line connecting the cpCas7 subunit proteins), the C-terminus of the second cpCas7 subunit protein is connected with the N-terminus of a different protein sequence (FIG. 6C, FP) (e.g., a cytidine deaminase) using linker polypeptides (FIG. 6C, straight black lines connecting cpCas7 and FP), and the C-terminus of this protein coding sequence is connected with the N-terminus of the third cpCas7 using a linker polypeptide. One advantage of such a fused backbone of cpCas7 subunit proteins is that an additional protein sequence can be introduced at a specific location along the backbone to provide access of the additional protein sequence to different locations along the length of the nucleic acid target sequence to which the guide directs binding of the Cascade complex.

[00193] FIG. 7A and FIG. 7B illustrate further embodiments of engineered Type I CRISPR-Cas effector complexes comprising fusion proteins. In FIG. 7A and FIG. 7B, the relative positions of the Cas proteins of the Cascade complex are shown (FIG. 7A, FIG. 7B: Cas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin). FIG. 7A shows a Cascade complex comprising a Cse2-Cse2 fusion protein (FIG. 7A, two Cse2 proteins connected by a curved black line). In silico design, cloning, expression, purification, and electrophoretic mobility shift assays are described in Example 11B and Example 11C Cascade complexes comprising Cse2-Cse2 fusion proteins. FIG. 7B shows a Cascade complex comprising a Cse2-Cse2 fusion protein connected via a linker polypeptide (FIG. 7B, curved black line connecting Cse2 protein to FP) with an additional protein sequence (FIG. 7B, FP). Example 11D describes in silico design, cloning, expression, and purification of a Cse2-Cse2 protein fused to a cytidine deaminase.

[00194] In some embodiments, one or more nuclear localization signals can be added at the engineered N-terminus or C-terminus of a Cascade protein subunit (e.g., a Cas8-FokI fusion protein, a cpCas7 protein, or a Cse2-Cse2 fusion protein).

[00195] In some embodiments of fusion polypeptides, linker polypeptides connect two or more protein coding sequences. The length of exemplary linker polypeptides is described in the Examples. Typically, linker lengths include, but are not limited to, between about 10 amino acids and about 40 amino acids, between about 15 amino acids and about 30 amino acids, and between about 17 amino acids and about 20 amino acids. The amino acid composition of linker polypeptides typically comprises amino acids that are polar, small, and/or charged (e.g., Gly, Ala, Leu, Val, Gln, Ser, Thr, Pro, Glu, Asp, Lys, Arg, His, Asn, Cys, Tyr). In additional embodiments, linker polypeptides are designed such that they do not contain methionine, and the fusion is designed to avoid cryptic translation initiation sites. Following the guidance of the present Specification, the linker polypeptide is designed to provide appropriate spacing and positioning of the functional domain and the Cascade protein within the fusion protein (see, e.g., Chichili, C., et al., Protein Science 22:153-167 (2013); Chen, X., et al., 65:1357-1369 (2013); George, R., et al., Protein Engineering, Design and Selection 15:871–879 (2002)). Additional examples of linker polypeptides useful in the practice of the present invention are linker polypeptides identified that connect coding sequences of Cascade proteins to each other in organisms comprising Cascade systems (e.g., the linker polypeptide that connects Cas8 to Cas3 in *Streptomyces griseus* as described by Westra, E.R., et al., Mol, Cell. 46: 595–605 (2012)).

[00196] Fusion protein coding DNA sequences can be codon-optimized for expression in a selected organism such as bacteria, archae, plants, fungi, or mammalian cells. Codon-optimizing programs are widely available, such as on the Integrated DNA Technologies website (www.idtdna.com/CodonOpt), or through Genscript® (Genscript, Piscataway, NJ) services. To facilitate cloning into the recipient expression vector, additional sequences overlapping with the vector compatible for SLIC cloning (see, e.g., Li, M., et al., Methods Mol. Biol. 852:51-59 (2012)) can be appended at the 5' and 3' ends of the DNA sequence.

[00197] In other embodiments, Cascade subunit proteins can be fused to transcription activation and/or repression domains. In some embodiments, a fusion protein can comprise activator domains (e.g., heat shock transcription factors, NFKB activators, VP16, and VP64 (see, e.g., Eguchi, A. et. al., Proc. Natl. Acad. Sci. USA 113:E8257-E8266 (2016); Perez-Pinera, P. et. al., Nature Methods 10:973-6 (2013); Gilbert, L.A., et. al. Cell 159:647-61 (2014)) or repressor domains (e.g., a KRAB domain). In some embodiments, linker nucleic acid sequences are used to join the two or more coding sequences for proteins, protein domains, or protein fragments.

[00198] Cascade complexes comprising Type I CRISPR-Cas subunit proteins fused to transcription activators can be used to activate the expression of the gene. The target locus can contain a transcriptional start site (TSS) that typically harbors one or more binding site for the transcriptional activation machinery (factors) of a cell. FIG. 8 illustrates a Cascade complex comprising six fusion proteins comprising a cpCas7 (compare to FIG. 3A) connected via a linker polypeptide (FIG. 8, curved black line connecting cpCas7 to VP64) to the transcriptional activator VP64. In FIG. 8, crRNA is illustrated as a dark black line comprising a hairpin, and the relative positions of the Cas proteins of the Cascade complex are shown (FIG. 8: cpCas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin). Such engineering of a Cascade complex converts the complex into a flexible tool for transcriptional activation of a gene (CASCADEa), wherein targeting a selected gene is achieved by selection of a guide sequence that directs binding of the Cascade complex to one or more regulatory elements (e.g., a TSS) of the selected gene. Example 12 describes the design of a *E. coli* Type I-E cp-Cas7 protein fused to a VP64 activation domain to confer transcriptional activation activity to the Cascade complex. Transcription activators include, but are not limited to, homeodomain proteins, zinc-finger proteins, winged-helix (forkhead) proteins, leucine-zipper proteins, helix-loop-helix proteins, heterodimeric transcription factors, activation domains, and transcription factors that bind enhancers (see, e.g., Molecular Cell Biology, Harvey Lodish, et al., W H Freeman & Co; (2002) ISBN 978-0849394805).

[00199] In addition, Cascade complexes comprising Type I CRISPR-Cas subunit proteins fused to transcription repressors can be used to repress the expression of the gene. The target locus can comprise transcriptional regulatory elements. In one embodiment, a Cascade subunit protein can be connected to a KRAB domain via a linker polypeptide. A Cascade complex comprising the Cascade subunit protein/KRAB domain fusion can convert the complex into a flexible tool for transcriptional repression of a gene (CASCADEi), wherein targeting a selected gene is achieved by selection of a guide sequence that directs binding of the Cascade complex to one or more regulatory elements of the selected gene. Transcriptional repressors include, but are not limited to, passive transcriptional repressors, bzip transcription factor family, sp1-like transcriptional repressors, active transcriptional repressors (e.g., transcriptional repression via recruitment of histone deacetylases, histone deacetylation, and dual-specific repressors (see, e.g., Thiel, G., et al., Eur. J. Biochem. 271:2855–2862 (2004); Nicola Reynolds, N., et al., Development 140:505-512 (2013); Gaston, K., et al., Cell Mol. Life Sci., 60:721-741 (2003)).

[00200] In additional embodiments, Cascade subunit proteins can be fused to affinity tags.

[00201] In other embodiments of the present invention, Type I CRISPR-Cas guide polynucleotides can be modified by insertion of a selected polynucleotide element or changes of a nucleotides at selected positions within the guide polynucleotides (e.g., a fundamentally different change of a DNA moiety for an RNA moiety, as well as other changes described above for guide polynucleotides). Such embodiments include, but are not limited to, Type I CRISPR-Cas guide polynucleotides 5', 3', or internally fused to one or more nucleotide effector domain (e.g., an MS2 or MS2-P65-HSF1 binding RNA or aptamer that recruits transcription factors). FIG. 9 illustrates a Type I CRISPR guide polynucleotide and the relative positions of the Cas proteins of the Cascade complex are shown (FIG. 9: Cas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin within the dashed box). In FIG. 9, the crRNA further comprises an RNA aptamer hairpin (FIG. 9, position indicated by the arrow) introduced into the 3' hairpin of the guide polynucleotide.

[00202] The length of Type I CRISPR-Cas guides can also be modified, typically by lengthening or shortening the Cas7 subunit protein and Cse2 subunit protein binding region. FIG. 10A illustrates a Cascade complex with three Cas7 subunits, one Cse2 subunit, and a shortened crRNA (FIG. 10A: Cas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line

comprising the hairpin). FIG. 10B illustrates a Cascade complex with nine Cas7 subunits, three Cse2 subunits, and a lengthened crRNA (FIG. 10B: Cas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin).

[00203] Example 16 describes the generation and testing of modifications of Type I CRISPR-Cas guide crRNAs and the suitability of the modified guides for use in constructing engineered Type I CRISPR-Cas effector complexes.

[00204] In a third aspect, the present invention relates to nucleic acid sequences encoding one or more engineered Cascade components, as well as expression cassettes, vectors, and recombinant cells comprising nucleic acid sequences encoding one or more engineered Cascade components. Some embodiments of the third aspect of the invention include one or more polypeptide encoding all the components of a selected Cascade system (e.g., Cse2, Cas5, Cas6, Cas7, and Cas8 proteins, and one or more cognate guides), wherein the components are capable of forming an effector complex. Typically, when more than one cognate guide is expressed, the guides have different spacer sequences to direct binding to different nucleic acid target sequences. Such embodiments include, but are not limited to, expression cassettes, vectors, and recombinant cells.

[00205] In one embodiment, the present invention relates to one or more expression cassettes comprising one or more nucleic acid sequences encoding one or more engineered Cascade components. Expression cassettes typically comprise a regulatory sequence involved in one or more of the following: regulation of transcription, post-transcriptional regulation, or regulation of translation. Expression cassettes can be introduced into a wide variety of organisms including, but not limited to, bacterial cells, yeast cells, plant cells, and mammalian cells (including human cells). Expression cassettes typically comprise functional regulatory sequences corresponding to the organism(s) into which they are being introduced.

[00206] A further embodiment of the present invention relates to vectors, including expression vectors, comprising one or more nucleic acid sequences encoding one or more one or more engineered Cascade components. Vectors can also include sequences encoding selectable or screenable markers. Furthermore, nuclear targeting sequences can also be added, for example, to Cascade subunit proteins. Vectors can also include polynucleotides encoding protein tags (e.g., poly-His tags, hemagglutinin tags, fluorescent protein tags, and bioluminescent tags). The coding sequences for such protein tags can be fused to, for example, one or more nucleic acid sequences encoding a Cascade subunit protein.

[00207] General methods for construction of expression vectors are known in the art. Expression vectors for host cells are commercially available. There are several commercial software products designed to facilitate selection of appropriate vectors and construction thereof, such as insect cell vectors for insect cell transformation and gene expression in insect cells, bacterial plasmids for bacterial transformation and gene expression in bacterial cells, yeast plasmids for cell transformation and gene expression in yeast and other fungi, mammalian vectors for mammalian cell transformation and gene expression in mammalian cells or mammals, and viral vectors (including, but not limited to, lentivirus, retrovirus, adenovirus, herpes simplex virus I or II, parvovirus, reticuloendotheliosis virus, and adeno-associated virus (AAV) vectors) for cell transformation and gene expression and methods to easily allow cloning of such polynucleotides.

[00208] AAV-based vectors (rAAV) are one example of viral vectors useful in the practice of methods of the present invention. AAV is a single-strand DNA member of the family Parvoviridae, and is a naturally replication-deficient virus. AAV vectors are among the viral vectors most frequently used for gene therapy. Twelve human serotypes of AAV (AAV serotype 1 [AAV-1] to AAV-12) and more than 100 serotypes from non-human are known. In one embodiment, AAV-6 is used as a vector.

[00209] Lentiviral vectors are another example of viral vectors useful in the practice of methods of the present invention. Lentivirus is a member of the Retroviridae family and is a single-stranded RNA virus, which can infect both dividing and non-dividing cells as well as provide stable expression through integration into the genome. To increase the safety of lentiviral vectors, components necessary to produce a viral vector are split across multiple plasmids. Transfer vectors are typically replication incompetent and may additionally contain a deletion in the 3'LTR, which renders the virus self-inactivating after integration. Packaging and envelope plasmids are typically used in combination with a transfer vector. For example, a packaging plasmid can encode combinations of the Gag, Pol, Rev, and Tat genes. A transfer plasmid can comprise viral LTRs and the psi packaging signal. The envelope plasmid usually comprises an envelope protein (usually vesicular stomatitis virus glycoprotein, VSV-GP, because of its wide infectivity range).

[00210] Illustrative plant transformation vectors include those derived from a Ti plasmid of *Agrobacterium tumefaciens* (see, e.g., Lee, L.Y., et al., Plant Physiology 146:325-332 (2008)). Also, useful and known in the art are *Agrobacterium rhizogenes* plasmids. For example, SNAPGENE™ (GSL Biotech LLC, Chicago, IL; snapgene.com/resources/plasmid_files/ your_time_is_valuable/) provides an extensive list of

vectors, individual vector sequences, and vector maps, as well as commercial sources for many of the vectors.

[00211] In order to express and purify recombinant Cascade in a bacterial expression system, vectors can be designed that encode Cascade subunit proteins, as well as a minimal CRISPR arrays comprising guide sequences of interest. Accordingly, one aspect of the present invention includes such expression systems. In one embodiment, the Cascade complex is expressed off of three distinct plasmid vectors, which collectively encode the following components: a Cas8 protein; Cse2, Cas7, Cas5, and Cas6 proteins; and a CRISPR RNA. In some embodiments, the expression plasmid encoding Cas8 comprises the natural gDNA gene sequence and, in other embodiments, the expression plasmid can encode Cas8 that is codon-optimized for expression in a chosen cell type. Similarly, the expression plasmid encoding Cse2, Cas7, Cas5, and Cas6 can contain the natural gDNA gene sequences or can contain gene sequences that have been codon-optimized for expression in a chosen cell type. In some embodiments, the entire Cascade subunit protein coding operon can be placed downstream of a single transcriptional promoter, such that the different proteins are all translated from a single polycistronic transcript. In additional embodiments, the gene encoding the Cascade subunit proteins can be separated from each other, with intervening transcriptional terminators and promoters.

[00212] The expression plasmid encoding the crRNA may contain as few as two repeats flanking a single spacer sequence, downstream of an appropriate transcriptional promoter, or may contain many repeats flanking multiple spacer sequences, of either the same exact guide sequence or multiple distinct guide sequences. Coordinated expression of the CRISPR and the Cascade subunits, in particular the Cas6 subunit, lead to processing of long precursor crRNAs into the mature length crRNA, each one of which comprises fragments of a single repeat on the 5' and 3' ends of the crRNA, and a single spacer sequence in the middle.

[00213] An alternative strategy to express the complete Cascade complex in *E. coli* uses two plasmids: one plasmid that encodes the entire Cas8–Cse2–Cas7–Cas5–Cas6 operon on a single expression plasmid and one plasmid that encodes the CRISPR RNA. In this case, the 5' end of the Cse2 gene, which normally overlaps with the 3' end of the Cas8 gene, is separated spatially from the 3' end of the Cas8 gene, in order to append a polynucleotide sequence encoding an affinity tag and/or protease recognition sequence.

[00214] Example 2 describes two types of bacterial expression plasmid systems for the Cascade proteins: the first type comprises two plasmids, a first plasmid encoding the Cas8 protein and a second encoding the 4 subunit proteins of the CasBCDE complex (*cse2–cas7–*

cas5–cas6 operon); and the second type comprises an expression plasmid encoding all five subunit proteins of the Cascade complex (*cas8–cse2–cas7–cas5–cas6* operon). Cognate CRISPR arrays are also described.

[00215] In order to facilitate purification of Cascade complexes, an affinity tag can be appended onto the Cse2 subunit, such as an N-terminal Strep-II tag or a hexahistidine (His6) tag. Furthermore, an amino acid sequence recognized by a protease, such as TEV protease or the HRV3C protease can be inserted between the affinity tag and the native N-terminus of the Cse2 subunit, such that biochemical cleavage of the sequence with the protease after initial purification liberates the affinity tag from the final recombinant Cascade complex. The affinity tag may also be placed on other subunits, or left on the Cse2 subunit and combined with additional affinity tags on other subunits. Exemplary Cascade subunit proteins comprising affinity tags are set forth in Example 1, Example 2, Example 3A, Example 3B, and Example 3C.

[00216] For Type I-E Cascade systems, a strain of *E. coli* can be transformed with plasmids encoding the CRISPR RNA as well as the *cse2–cas7–cas5–cas6* genes, protein expression induced, and a Cascade complex that is lacking the Cas8 subunit can be produced. This Cascade complex typically is referred to as a Cas8-minus Cascade complex, or alternatively as a CasBCDE complex (see, e.g., Jore, M., et al., Nat. Struct. Mol. Biol. 18:529-536 (2011)). This purified complex can be biochemically combined with separately purified Cas8 to reconstitute full Cascade (see, e.g., Sashital, D.G., et al., Mol. Cell 46:606-615 (2012)).

[00217] Table 6 presents exemplary sequences of bacterial expression plasmids encoding the minimal CRISPR array, *cas8*, *cse2–cas7–cas5–cas6* constructs, and *cas8–cse2–cas7–cas5–cas6* constructs, containing different tags and designs. Plasmids that encode Cascade complexes and Cascade complexes from homologous Type I systems can be designed similarly as the exemplary expression plasmid sequences for the Type I-E found in *E. coli* K-12 MG1655 following the guidance of the present Specification. Table 6 additionally contains sequences of expression plasmids expressing Cas8–Cse2–Cas7–Cas5–Cas6 proteins as well as FokI fusions to either the *cas8* gene or the *cas6* gene, for the production of nuclease–Cascade fusions for gene editing experiments.

[00218]

Table 6 Vectors for Production of Cascade Effector Complexes			
SEQ ID NO:	**Description**	**Effector complex species of origin**	**Type of sequence**
SEQ ID NO:352	minimal CRISPR array	I-E_*E. coli* K-12 MG1655	Spacer sequence targets J3
SEQ ID NO:353	minimal CRISPR array	I-E_*E. coli* K-12 MG1655	Spacer sequence targets CCR5.1
SEQ ID NO:354	minimal CRISPR array (J3/L3)	I-E_*E. coli* K-12 MG1655	Minimal CRIPSR array spacers sequence targets J3 and L3
SEQ ID NO:355	minimal CRISPR array (Hsa07)	I-E_*E. coli* K-12 MG1655	Minimal CRISPR array spacers sequence targets Hsa07
SEQ ID NO:356	His6-MBP-TEV-Cas8	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:357	StrepII-HRV3C-Cse2_Cas7_Cas5_Cas6	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:358	Cas8_His6–HRV3C–Cse2_Cas7_Cas5_Cas6	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:359	FokI–30aa–Cas8_His6–HRV3C–Cse2_Cas7_Cas5_Cas6	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:360	FokI–30aa–Cas8_His6–HRV3C–Cse2_Cas7_Cas5_NLS–Cas6	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:361	FokI–30aa–Cas8_His6–HRV3C–Cse2_Cas7–NLS_Cas5_Cas6	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:362	Cas8_His6–HRV3C–Cse2_Cas7_Cas5_Cas6–20aa–FokI	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags

[00219] Table 7 contains the sequences of single polypromoter bacterial expression plasmids encoding all five subunit proteins together with the crRNA from a single bacterial expression plasmid. In this design, each gene is separated from the other genes it flanks upstream and downstream with a transcriptional promoter and terminator. Additional sequences can be introduced that encode an affinity tag and/or protease recognition tag, as well as a fusion to a nuclease protein, in order to generate a Cascade–nuclease fusion for gene editing.

[00220]

Table 7 Vectors for Production of Cascade Effector Complexes			
SEQ ID NO:	**Description**	**Effector complex species of origin**	**Type of sequence**
SEQ ID NO:363	Polypromoter, Cas5_Cas3_Cse2_ Cas7_Cas6_Cas8_CRISPR(J3)	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:364	Polypromoter, Cas5_Cas3_Cse2_ Cas7_CRISPR(J3)_Cas6_Cas8	I-E_*E. coli* K-12 MG1655	Derived from gDNA, with appended tags
SEQ ID NO:365	Polypromoter(EcoCO), CRISPR(J3/L3)_ Cse2_Cas7_Cas5_Cas8_Cas6	I-E_*E. coli* K-12 MG1655	*E. coli* codon-optimized DNA gene sequences
SEQ ID NO:366	Polypromoter(EcoCO), CRISPR(J3/L3)_Cse2_Cas7_Cas5_ Cas8_FokI-30aa-Cas6	I-E_*E. coli* K-12 MG1655	*E. coli* codon-optimized DNA gene sequences
SEQ ID NO:367	Polypromoter(EcoCO), CRISPR(J3/L3)_Cse2_Cas7_Cas5_ Cas6_FokI-30aa-Cas8	I-E_*E. coli* K-12 MG1655	*E. coli* codon-optimized DNA gene sequences

[00221] Additional bacterial expression plasmids can be designed encoding homologous Cascade complexes from other Type I subtypes and other bacterial or archaeal organisms based on the design criteria herein. Such expression plasmids can be designed with gDNA sequences for the Cascade genes, or they can be designed with gene sequences that have been codon-optimized for expression in *E. coli* or other bacterial strains.

[00222] In order to express Cascade or effectors fusions to Cascade in mammalian cells, such as human cells, eukaryotic expression plasmid vectors were designed to enable expression of the relevant proteins and RNA components by eukaryotic transcription and translation machinery. In one embodiment, Cascade can be generated in mammalian cells by encoding each of the protein components on a separate expression vector driven by a eukaryotic promoter (e.g., a cytomegalovirus (CMV) promoter), and encoding the crRNA on a separate expression vector driving by an RNA polymerase III promoter (e.g., the human U6 promoter). The CRISPR RNA can be encoded with a minimal CRISPR array containing at least two repeats flanking one or more spacer sequences that function as the guide portion of the mature crRNA. The construct generating CRISPR RNA can be designed with additional sequences flanking the outermost repeats in the minimal array. Processing of the precursor CRISPR RNA is enabled by the RNA processing subunit of the Cascade complex (Cas6 subunit protein), which can be expressed from a separate plasmid.

[00223] Table 8 contains the sequences of individual eukaryotic expression plasmids for each protein of the *E. coli* Type I-E Cascade complex. Cas8 subunit can be fused to additional effector nuclease domains, such as the FokI nuclease (Example 1, Example 3A, Example 3B, and Example 3C). Table 8 also contains the sequences of expression plasmids for the crRNA component of Cascade, encoding two separate crRNAs, whereby three repeat sequences flank two spacer spacers. Each of the protein-coding genes can be appended to polynucleotide sequences that append nuclear localization signals (NLS), affinity tags, and linker sequences connecting those tags. Other fusions to any of the Cascade subunit proteins can be encoded by additional polynucleotide sequences that typically are appended to either the 5' or 3' coding sequence, including additional polynucleotide sequences that encode amino acid linkers connecting to the Cascade subunit protein to additional polypeptide sequences of interest. Examples of candidate fusions proteins are described herein.

[00224]

Table 8
Vectors for Production of Cascade Effector Complexes

SEQ ID NO:	Description	Effector complex species of origin	Type of sequence
SEQ ID NO:368	Cas8, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:369	NLS–Cas8, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:370	NLS–HA–FokI–30aa–Cas8, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:371	NLS–Cse2, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:372	NLS-Cas7, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:373	Cas5, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:374	NLS–Cas5, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:375	Cas6, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence

Table 8 Vectors for Production of Cascade Effector Complexes			
SEQ ID NO:	**Description**	**Effector complex species of origin**	**Type of sequence**
SEQ ID NO:376	NLS–Cas6, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:377	NLS–V5–FokI–30aa–Cas8, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:378	Cas3–NLS, HsCO	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:379	CRISPR(Hsa07)	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence

[00225] In order to express components of the Cascade complex on fewer expression vectors, polycistronic expression vectors can be constructed, whereby a single promoter (e.g., a CMV promoter) drives expression of multiple coding sequence simultaneously that are separated by a *Thosea asigna* virus 2A sequence. 2A viral peptide sequences induce ribosomal skipping, enabling multiple protein-coding genes to be concatenated within a single polycistronic construct for expression in eukaryotic cells. Thus, polycistronic vectors can be designed that encode four or five protein subunits of the Cascade complex on a single transcript driven by a single promoter. Table 9 contains the sequences of eukaryotic polycistronic expression plasmids that can be combined with a CRISPR RNA expression plasmid to produce functional Cascade in mammalian cells.

[00226]

Table 9 Vectors for Production of Cascade Effector Complexes			
SEQ ID NO:	**Description**	**Effector complex species of origin**	**Type of sequence**
SEQ ID NO:380	Polycistronic(HsCO), NLS–Cas7_NLS–Cse2_NLS–Cas5_NLS–Cas6	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:381	Polycistronic(HsCO), NLS–Cas7_NLS–Cse2_NLS–Cas5_NLS–Cas6_NLS–Cas8	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:382	Polycistronic(HsCO), NLS–Cas7_NLS–Cse2_NLS–Cas5_NLS–Cas6_NLS–FokI–30aa–Cas8	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence

Table 9 Vectors for Production of Cascade Effector Complexes			
SEQ ID NO:	**Description**	**Effector complex species of origin**	**Type of sequence**
SEQ ID NO:383	Polycistronic(HsCO), NLS–Cas7_NLS–Cse2_NLS–Cas5_NLS–Cas6_NLS–FokI–30aa–Cas8, no epitope tags	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:384	Polycistronic(HsCO), NLS–Cas7_NLS–Cse2_NLS–Cas5_NLS–FokI–30aa–Cas6_NLS–Cas8, no epitope tags	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence

[00227] In some embodiments, the CRISPR RNA is encoded within the 3' untranslated region (UTR) of a protein-coding gene, the expression of which is driven by an RNA polymerase II promoter (e.g., CMV promoter) to produce a transcript. In such embodiments, the minimal CRISPR array is designed to exist downstream of a protein coding gene such as Cas6, Cas7, or a reporter gene (e.g., an enhanced green fluorescent protein, eGFP), and is separated from the protein coding sequence by a MALAT1 triplex sequence that has previously been shown to confer stability to the upstream transcript. The minimal CRISPR array is processed by the RNA processing subunit of Cascade (typically expressed using a different plasmid), an endonuclease that cleaves the minimal CRISPR array, a break is introduced into the transcript, and the triplex sequence protects the 3' end of the upstream protein-coding gene from premature exonucleolytic degradation. Table 10 contains sequences of three polynucleotide sequences, whereby the CRISPR array is cloned downstream of either Cas6, Cas7, or eGFP, and expression of the entire fusion sequence is driven by a CMV promoter.

[00228]

Table 10 Vectors for Production of Minimal CRISPR Arrays			
SEQ ID NO:	**Description**	**Effector complex species of origin**	**Type of sequence**
SEQ ID NO:385	eGFP_MALAT1-triplex_CRISPR (Hsa07)	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:386	NLS–Cas7_MALAT1-triplex_CRISPR (Hsa07)	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:387	NLS–Cas6_MALAT1-triplex_CRISPR (Hsa07)	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence

[00229] In some embodiments, the CRISPR RNA array is encoded on the same vector as the polycistronic construct driving expression of the five5 Cascade subunit proteins; the combination of these two elements generates an all-in-one vector that produces all functional subunits (both protein and RNA) of the Cascade complex, together with any nuclease or effector domains fused to one of the Cascade subunits. Table 11 contains two representative sequences of these all-in-one polynucleotide sequences that encode all the respective components to produce functional FokI-Cascade RNPs in mammalian cells.

[00230]

Table 11
Vectors for Production of Cascade Effector Complexes

SEQ ID NO:	Description	Effector complex species of origin	Type of sequence
SEQ ID NO:388	hU6_CRISPR(Hsa07)_F, CMV_NLS–Cas7_NLS–Cse2_NLS–Cas5_NLS–Cas6_NLS–FokI–30aa–Cas8	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:389	hU6_CRISPR(Hsa07)_R, CMV_NLS–Cas7_NLS–Cse2_NLS–Cas5_NLS–Cas6_NLS–FokI–30aa–Cas8	I-E_*E. coli* K-12 MG1655	*H. sapiens* codon-optimized DNA gene sequence

[00231] Example 3A, Example 3B, and Example 3C describe expression systems using separate plasmids expressing each Cascade subunit protein and minimal CRISPR array, expression systems wherein multiple Cascade subunit protein coding sequences are expressed from a single promoter, and an expression system wherein a single-plasmid Cascade expression system was constructed to express the entire *cas8–cse2–cas7–cas5–cas6* operon and a minimal CRISPR array for use in mammalian cells.

[00232] One of ordinary skill in the art following the guidance of the present Specification can design additional mammalian expression vectors encoding other Cascade complexes analogously to the examples provided the *E. coli* Type I-E Cascade complex.

[00233] In a fourth aspect, the present invention relates to production of engineered Type I CRISPR-Cas effector complexes by introduction of plasmids encoding one or more components of the engineered Type I CRISPR-Cas effector complexes into host cells. Transformed host cells (or recombinant cells) or the progeny of cells that have been transformed or transfected using recombinant DNA techniques can comprise one or more nucleic acid sequences encoding one or more component of an engineered Type I CRISPR-Cas effector complex. Methods of introducing polynucleotides (e.g., an expression vector) into host

cells are known in the art and are typically selected based on the kind of host cell. Such methods include, for example, viral or bacteriophage infection, transfection, conjugation, electroporation, calcium phosphate precipitation, polyethyleneimine-mediated transfection, DEAE-dextran mediated transfection, protoplast fusion, lipofection, liposome-mediated transfection, particle gun technology, microprojectile bombardment, direct microinjection, and nanoparticle-mediated delivery. In one embodiment of the present invention, polynucleotides encoding components of engineered Type I CRISPR-Cas effector complexes are introduced into bacterial cells (e.g., *E. coli*).

[00234] Example 4A and Example 4B describe a method for introduction and expression of Cas8 protein coding sequences, as well as coding sequences for components of engineered Type I CRISPR-Cas effector complexes for bacterial production of such complexes using *E. coli* expression systems.

[00235] A variety of exemplary host cells disclosed herein can be used to produce recombinant cells using an engineered Cascade effector complex. Such host cells include, but are not limited to, a plant cell, a yeast cell, a bacterial cell, an insect cell, an algal cell, and a mammalian cell.

[00236] For ease of discussion, "transfection" is used below to refer to any method of introducing polynucleotides into a host cell.

[00237] In some embodiments, a host cell is transiently or non-transiently transfected with nucleic acid sequences encoding one or more component of a Type I CRISPR-Cas effector complex. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is first removed from a subject, e.g., a primary cell or progenitor cell. In some embodiments, the primary cell or progenitor cell is cultured and/or is returned after *ex vivo* transfection to the same subject or to a different subject.

[00238] Expression and purification of engineered Type I CRISPR-Cas effector complexes is labor intensive, so to facilitate screening across a large number of guide polynucleotide or effector complex variants, a higher throughput plasmid-based delivery system was designed. Each of the five *Cas* genes was human codon-optimized and cloned into a CMV-driven expression plasmid as an N-terminal NLS fusion, and a minimal CRISPR array containing paired gRNAs targeting the TRAJ27 exon of the T cell receptor alpha locus (UCSC genome browser, hg38) was cloned into a sixth plasmid downstream of a human U6 promoter (Example 3A; FIG. 35). In FIG. 35, the order of the elements from left to right is as follows: hu6 promoter, grey rectangle with diamond end; repeat 1, open diamond, (white); spacer 1, grey waffle rectangle; repeat 2, grey diamond; spacer 2, grey stipple rectangle; and repeat 3,

black diamond. In FIG. 35, the bracket illustrates the region encoding two gRNAs. In some embodiments, the two guide RNAs can be the same (e.g., target the same nucleic acid target sequence), and in other embodiments the two guide RNAs can be different (e.g., target two different nucleic acid target sequences).

[00239] gRNA processing in most Type I systems is naturally catalyzed by the Cas6 ribonuclease present in Cascade (see, e.g., Brouns, S. J., et al., Science 321:960-964 (2008); Hochstrasser, M., et al., Trends Biochem. Sci. 40:58–66 (2015), obviating the need for multiple promoters with the paired gRNA approach set forth herein. Accordingly, one embodiment of the present invention comprises vectors comprising paired guide polynucleotides operably linked to regulatory elements to provide expression of the guide polynucleotides (e.g., gRNAs). Six-plasmid co-transfection yielded up to ~3% editing at the TRAJ27 locus, and removal of any one component abrogated genome editing, with the sole exception of Cas11, which the *E. coli* Cascade effector complex does not absolutely require for DNA binding (see, e.g., Westra, E., et al., RNA Biol. 9:1134-1138 (2012)).

[00240] In another embodiment of the present invention, minimal CRISPR arrays, typically comprising two guide sequences, are introduced into cells or biochemical reactions as DNA templates. The DNA templates are produced by PCR amplification (e.g., FIG. 42A; Example 20A). Such minimal CRISPR arrays can be introduced into cells with one or more plasmids encoding the Cascade complex protein components. In some embodiments, minimal CRISPR arrays and vectors comprising paired guide polynucleotides can both be introduced into a cell or biochemical reaction. In methods using two Cascade RNP complexes (e.g., methods of binding a nucleic acid target sequence or methods of cutting a nucleic acid target sequence; see, e.g., FIG. 15A, FIG. 15B, FIG. 15C), minimal CRISPR arrays can encode two different guides. Accordingly, in some embodiments the two guide RNAs can be different (e.g., target two different nucleic acid target sequences). In methods using a single Cascade RNP complex (e.g., when using one Type I CRISPR-Cas effector complex associated with a mCas3 protein or a Type I CRISPR-Cas effector complex wherein a Cas3 fusion protein associates with the complex; e.g., see, e.g., FIG. 16A, 17B, 17C, FIG. 21A, FIG. 21B, FIG. 21C, FIG. 21D), minimal CRISPR arrays can encode two copies of the same guide sequence. Accordingly, in some embodiments, the two guide RNAs can be the same (e.g., target the same nucleic acid target sequence).

[00241] In yet another embodiment, polynucleotides encoding guide sequences that further comprise sequences and structures recognized by Cas6 protein for the endonucleolytic processing of crRNA precursors to mature guide RNAs can be introduced into cells or

biochemical reactions. In other embodiments, mature guide polynucleotides that do not require processing can be used in assembly of Cascade complexes. Such mature guides can comprise sequence modifications (e.g., phosphorothioate linkages at the 5' and/or 3' ends to help protect the guide from nuclease digestion, such as by RNases). Additional guide modifications include those described herein for nucleotide sequences (e.g., nucleotide analogs, etc.).

[00242] Example 9A, Example 9B, Example 9C, and Example 9D illustrate the design and delivery of *E. coli* Type I-E Cascade complexes comprising FokI fusion proteins to facilitate genome editing in human cells. Example 9B describes the delivery of plasmid vectors expressing Cascade complex components into eukaryotic cells. In a fifth aspect, the present invention relates to the purification of engineered Type I CRISPR-Cas effector complexes from cells and uses of such complexes. Engineered Type I CRISPR-Cas effector complexes are produced in a host cell. The engineered Type I CRISPR-Cas effector complexes (in this case Cascade RNP complexes) are purified from cell lysates.

[00243] Example 5A and Example 5B describe purification of *E. coli* Type I-E Cascade RNP complexes produced by overexpression in bacteria as described in Example 4B. The method uses immobilized metal affinity chromatography followed by size exclusion chromatography (SEC). Example 5A and Example 5B describe methods that can be used to assess the quality of purified Cascade RNP products. Examples are presented illustrating the purification of Cas8, Cas7, Cas6, Cas5, and Cse2 Cascade RNP complexes, Cascade complexes comprising Cas7, Cas6, Cas5, and Cse2 proteins, and FokI-Cas8 fusion proteins.

[00244] The purified, engineered Type I CRISPR-Cas effector complexes can also be used directly in biochemical assays (e.g., binding and/or cleavage assays). Example 6A, Example 6B, and Example 6C describe production of dsDNA target sequences for use in *in vitro* DNA binding or cleavage assays. Example 6 describes three methods to produce target sequences, including annealing of synthetic ssDNA oligonucleotides, PCR amplification of selected nucleic acid target sequences from gDNA, as well as cloning of nucleic acid target sequences into bacterial plasmids. The dsDNA target sequences were used in Cascade binding or cleavage assays.

[00245] The site-specific binding of and/or cutting by one or more engineered Type I CRISPR-Cas effector complexes can be confirmed, if necessary, using an electrophoretic mobility shift assay (see, e.g., Garner, M., et al., Nucleic Acids Res. 9:3047-3060 (1981); Fried, M., et al., Nucleic Acids Res. 9:6505-6525 (1981); Fried, M., Electrophoresis 10:366-376 (1989); Fillebeen, C., et al., J. Vis. Exp. (94), e52230, doi:10.3791/52230 (2014)), or the biochemical cleavage assay described in Example 7.

[00246] The data presented in Example 7 demonstrate that engineered Type I CRISPR-Cas effector complexes can exhibit nearly quantitative DNA cleavage, as evidenced by conversion of a supercoiled, circular plasmid substrate into a cleaved, linear form. After demonstrating robust biochemical activity with engineered Type I CRISPR-Cas effector complexes (e.g., comprising FokI-Cascade component fusion proteins), genome editing in cells was performed.

[00247] Example 8A, Example 8B, Example 8C, and Example 8D illustrate the design and delivery of *E. coli* Type I-E Cascade complexes comprising Cas subunit protein-FokI fusion proteins to human cells. The data in Example 8D demonstrate delivery of pre-assembled Cascade RNPs into target cells and effective genome editing in human cells.

[00248] The purified, engineered Type I CRISPR-Cas effector complexes can be directly introduced into cells. Methods to introduce the components into a cell include electroporation, lipofection, particle gun technology, and microprojectile bombardment.

[00249] FIG. 36A, FIG. 36B, FIG. 36C, and FIG. 36D provide comparative data for genome editing in human cells using engineered Cascade-RNP complexes and plasmid-based delivery of engineered Type I CRISPR-Cas complexes. In FIG. 36A-D, FIG. 36A, HEK293 cells were transfected with purified RNPs followed by next-generation sequencing (NGS) analysis of edited sites. As is shown in FIG. 36A (RNP Transfection), FokI-Cascade RNP complexes (FIG. 36A, represented on the left side of the figure above the straight line) targeting two adjacent loci were nucleofected into HEK293 cells (FIG. 36A, star-shaped, grey, to the left of the figure) to induce DNA cleavage and genome editing. Editing efficiencies at 16 unique genomic target sites (see Example 6C, Table 31, Human Dual Hsa1-16) were calculated (n=1). TRAC is the constant region of the T cell receptor. When T cell receptors are generated, they include splice junctions (i.e., "variable" region and "joining" region). Some of the TRAC guides described herein target joining regions (e.g., TRAJ27). Interspacer distances for each target are shown below the graph (FIG. 36A, left to right, 25, 30, 35, 40, 45 base pairs (bp)). In FIG. 36A, the vertical axis is percent editing efficiency (FIG. 36A, Editing Efficiency (%)), the horizontal axis represents targets 1 to 16, and below the horizontal axis are brackets indicating the interspacer length in base pairs (bp).

[00250] FIG. 36B provides representative DNA repair outcomes for Target 7 in FIG. 36A. In FIG. 36B, the relative locations of the half-sites targeted by the paired gRNAs are shown at the top of the figure, with their associated PAM sites. The interspacer distance is illustrated by the top line. In the graph, the expected cleavage site (FIG. 36B, position "0" shown as vertical black mid-line) and bp distances (-50 to 50) are indicated at the top. Each

horizontal grey line represents a different class of sequenced reads that were observed at the targeted locus. Indicators for these lines are as follows: grey area=sequence match; horizontal black line=deletion; and open box=insertion. A circle is located to the right of the graph by each line: the black circle is a wild-type read; and the open white circles are mutant reads. The expected wild-type read is illustrated in the first grey bar ("Ref"; i.e., the reference sequence). The wild-type read is illustrated in the second grey bar (the second grey bar; FIG. 36B, black circle). The next 11 lines illustrate mutant reads (FIG. 36B, open circles). Insertion lengths, given in number of base pairs, are shown in the column to the right of the circles. The total percent of reads is shown in the next column to the right, and the total reads are presented in the last column to the right.

[00251] As shown in FIG. 36C (6-plasmid transfection system), HEK293 cells (FIG. 36C, star-shaped, grey, to the left of the figure) were transfected with six plasmids, five plasmids encoding a Cas proteins (FIG. 36C, plasmids indicated as FokI-Cas8, Cas11, Cas7, Cas5, and Cas6), and one plasmid encoding the paired gRNAs were under the control of a CMV and human U6 (hU6) promoter (FIG. 36C, gRNA), followed by NGS analysis of edited sites. Illustrations of the FokI-Cascade RNP complexes are below the dashed line. Editing efficiencies at Target 7 from FIG. 36A were calculated (n=2) (FIG. 36A, black bars in the graph), and plasmid mixtures lacking single components (FIG. 36C, below horizontal axis, grey boxes containing -/+) were included as controls (FIG. 36C, open bars in the graph).

[00252] As shown in FIG. 36D (2-plasmid transfection system), HEK293 cells (FIG. 36D, star-shaped, grey, to the left of the figure) were transfected with a paired gRNA expression plasmid (FIG. 36D, gRNA plasmid) and a polycistronic expression plasmid encoding all five proteins separated by T2A "ribosome skipping" sequences peptides (FIG. 36D, CMV-Cas7-2A-Cas11-2A-Cas5-2A-Cas6-2A-FokI-Cas8), followed by NGS analysis of edited sites. Illustrations of the FokI-Cascade RNP complexes are below the dashed line. Editing efficiencies at the 16 targets shown in FIG. 36A were calculated for both the 2-plasmid system transfections (FIG. 36D, open bars) and the 6-plasmid system transfections from FIG.37C (n=3) (FIG. 36D, black bars). In FIG. 36D, the vertical axis is percent editing efficiency ("Editing Efficiency (%)), the horizontal axis represents targets 1 to 16, and below the horizontal axis are brackets indicating the interspacer length in base pairs (bp) (FIG. 36D, left to right, 25, 30, 35, 40, 45 bp).

[00253] Experiments were carried out by nucleofecting HEK293 cells with purified Cascade-RNPs containing nuclear localization signal sequences on FokI and Cas6. Up to ~4% editing efficiency was observed, as evidenced by next-generation sequencing of PCR amplicons

obtained from gDNA and, among the 16 target sites tested, editing was typically at sites containing 30 bp interspacer lengths (FIG. 36A). Closer inspection of the spectrum of repair outcomes revealed that indels were clustered in the middle of the interspacer (FIG. 36B) consistent with the design of the Type I CRISPR-Cas complexes. Accordingly, in one embodiment of the present invention, the engineered Type I CRISPR-Cas complexes are introduced directly into a cell. For the 6-plasmid delivery experiments (FIG. 36C), plasmid mixtures were assembled containing 420 ng of each plasmid except one, and then either water as a negative control or 700 ng of the missing plasmid was added back subsequent to nucleofection. For the initial FokI-EcoCascade polycistronic 2-plasmid delivery experiments (FIG. 36D), cells were electroporated with 500 ng of each plasmid or 500 ng of paired gRNA expression plasmid and 2.5 μg of polycistronic plasmid (3 μg total for each condition). In one embodiment, all five *cas* genes were constructed into a single polycistronic expression vector (FIG. 36D) connected in series by T2A "ribosome skipping" sequences (see, e.g., Kim, J., et al., PLoS ONE 6, e18556 (2011); Liu, Z., et al., Sci. Rep. 7:2193 (2017)). Strikingly, co-transfection with the polycistronic plasmid and paired gRNA expression plasmid resulted in editing efficiencies and DNA repair outcomes similar to those observed with both the 6-plasmid method (Example 9A) and direct RNP delivery methods (Example 8A, Example 8B, Example 8C, Example 8D), supporting the conclusion that biochemically active engineered Type I CRISPR-Cas effector complexes were being assembled and trafficked to the nucleus in human cells. Collectively, these experiments validated a greatly simplified expression system to reconstitute an elaborate, 11-subunit RNA-guided nuclease in eukaryotic cells with just two molecular components that are similar in size to the widely used Cas9 and sgRNA plasmids.

[00254] The data for engineered Type I CRISPR-Cas complexes (*E. coli* (EcoCascade, *Pseudomonas sp. S-6-2* (PseCascade), and *Streptococcus thermophilus* (SthCascade)) suggested that most target sites would be unique since they must include both half-sites, the requisite interspacer distance, and permissive PAMs. Engineered Cascade homologs from EcoCascade, PseCascade, and SthCascade were selected for more detailed characterization.

[00255] FIG. 37A, FIG. 37B, FIG. 37C, and FIG. 37D, illustrate editing efficiency as related to the FokI linker, interspacer length, and Cascade homolog. FIG. 37A, FokI-EcoCascade editing efficiency is shown as a function of FokI-Cas8 linker length (FIG. 37A, open circles, low line 10 aa; open circle upper graph line, 20 aa; black circles, 17 aa; and grey circles, 30 aa linker lengths) and interspacer distance. In FIG. 37A the vertical axis is editing efficiency (%), and the horizontal axis is interspacer distance in bp. Each data point represents the average of 3–4 unique target sites.

[00256] FIG. 37B provides FokI-Cascade nucleases with 30-aa linkers. FokI-Cas8 linkers were generated for 12 Type I-E Cascade variants and tested for genome editing at 4-7 target sites. Each data point represents a single genomic site, and bars show the mean and standard deviation (s.d.) across sites. Targets contained either AAG (FIG. 37B, grey bars) or GAA (FIG. 37B, white bars) PAM sequences and 30 bp interspacer distances, wherein the species are on the horizontal axis as follows: Eco, *E. coli*; Pse, *Pseudomonas sp.* S-6-2; Sen, *Salmonella enterica*; Geo, *Geothermobacter sp.* EPR-M; Mar, *Methanocella arvoryzae*; Ahe, *Atlantibacter hermannii*; Oce, *Oceanicola sp.* HL-35; Pae, *Pseudomonas aeruginosa*; Sth, *Streptococcus thermophilus*; Str, *Streptomyces sp.* S4; Kpn, *Klebsiella pneumoniae*; Lba, *Lachnospiraceae bacterium*.

[00257] In FIG. 37C, FokI-PseCascade data is presented, wherein the vertical axis is percent editing efficiency (FIG. 37C, Editing Efficiency (%)) and the horizontal axis represents the interspacer length in base pairs (bp). The FokI-Cas8 linker length was 17 amino acids. Each data point represents a single genomic site, and bars show the mean and s.d. across 7–8 sites.

[00258] FIG. 37D provides data for FokI-PseCascade editing efficiency as a function of PAM sequence, the vertical axis is percent editing efficiency (FIG. 37D, Editing Efficiency (%)), and the horizontal axis corresponds to PAM sequences (FIG. 37 D, left to right, CCG, CGC, AAG, AAA, ATG, AAC, AGG, ATA, GAG, and AAT). Genomic sites contained one AAG PAM and a variable PAM at the second half-site, as shown on the horizontal axis. Each data point represents a single genomic site, and bars show the mean and s.d. across 6–15 sites.

[00259] FIG. 37E provides data for FokI-EcoCascade editing efficiency (FIG. 37E, vertical axis, editing efficiency (%)) as a function of PAM sequence. Target sites contained a fixed AAG PAM and a variable PAM at the second half-site, as shown on the horizontal axis (FIG. 37E, left to right, CCG, CGC, AAG, AGG, ATG, GAG, AAA, AAC, ATA, and AAT). Each dot represents a single target site in HEK293 cells and 6-15 sites were tested per PAM (n=1 per site). The bar graph displays the mean and s.d.

[00260] FIG. 37F provides data for FokI-SthCascade efficiency (FIG. 37F, vertical axis, editing efficiency (%)) as a function of PAM sequence. Target sites contained a fixed GAA PAM and a variable PAM at the second half-site, as shown on the horizontal axis (FIG. 37F, left to right, CC, AA, GA, TA, and CA). Each dot represents a single target site in HEK293 cells and 18-33 sites were tested per PAM (n=1 per site). The bar graph displays the mean and s.d.

[00261] FIG. 37G provides heat maps depicting the indel class frequencies for 40 genomic sites exhibiting high editing efficiencies (10–53%) from FIG. 37C and FIG. 37D.

Percent editing efficiency from 0-60 is presented in the bar graph in the top panel. Insertion lengths from 1–8 bp are presented in the heat map presented in the middle panel, and deletions lengths from 1–50 bp are presented in the heat map in the bottom panel. The 40 genomic target sites (FIG. 37G, Targets) are indicated on the horizontal axis (1-40). Single bp insertions are separated by nucleotide identity, and the grey scale intensity scales at the bottom of the figure correspond to insertion frequency percentage (FIG. 37G, Ins Freq (%), scale is 0 to greater than or equal to 20) and deletion frequency percentage (FIG. 37G, Del Freq (%), scale is 0 to greater than or equal to 20). The bar graph to the right displays the mean frequency (FIG. 37G, scale is 0 to 20) of each indel class. The pie chart to the right shows the fraction of 2–4 bp insertions resulting from putative templated repair (FIG. 37 G, black area of pie chart), defined here as containing a duplication of sequences adjacent to the cleavage site. "Other" is represented in grey area of pie chart.

[00262] The most closely related sites in the human genome for the five most highly edited FokI-PseCascade target sites (~20-48 % editing) were investigated, constrained only by a 30-33 bp interspacer requirement. Across all five targets, no sites with <22 mismatches across both half-sites were identified. For FokI-EcoCascade FokI-Cas8 linker type and interspacer distance experiments (FIG. 37A), cells were nucleofected with 2.4 μg of FokI-EcoCascade polycistronic plasmid and ~0.5-3.5 μg of paired gRNA expression plasmid.

[00263] For the FokI-Cascade homolog screen (FIG. 37B), cells were nucleofected with 1.5 μg of FokI-Cascade polycistronic plasmid and ~ 0.4-2.2 μg of paired gRNA expression plasmid. Across homologs, 4-7 sites were targeted, and sites were selected that showed high editing efficiency with FokI-EcoCascade. For the homolog variant FokI-Cas8 linker type and interspacer distance editing experiments (FIG. 37C and FIG. 41A to FIG. 41C), cells were nucleofected with 5 μg of polycistronic plasmid and ~100-400 ng of oligo-templated paired gRNA expression amplicon. For this experiment, gRNA concentrations were not normalized across wells or homolog variants. Additionally, for FIG. 41A to FIG. 41C, cells were nucleofected with, on average, ~1.5x more FokI-PseCascade gRNA than FokI-EcoCascade or FokI-SthCascade gRNA.

[00264] Oligo-templated PCR amplification is described herein (e.g., Example 20A). The oligo-templated PCR strategy to generate amplicons for paired gRNA expression from a human U6 (hU6) promoter (FIG. 42A, 420) in mammalian cells is illustrated in FIG. 42A and FIG. 42B. Briefly, the reverse inner oligonucleotide (FIG. 42A, 424) encodes both gRNA sequences and is modified for new target sites (also referred to as a unique primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence (FIG. 42A, 421: repeat, open rectangle; spacer 1,

grey rectangle; repeat, open rectangle; spacer 2, grey rectangle; repeat, open rectangle), whereas the remaining primers are invariant (FIG. 42A: forward outer primer, 422; forward inner primer, 423; reverse outer primer, 425). Editing efficiencies at target 7 (see FIG. 36B) after co-transfecting HEK293 cells with the polycistronic plasmid encoding FokI EcoCascade RNP complex and either a paired gRNA expression plasmid or paired gRNA expression amplicon are presented in FIG. 42B. In FIG. 42B, the vertical axis is editing efficiency (%) and the horizontal axis is paired gRNA cassette (ng). The data points are as follows: FokI-EcoCascade RNP complex (ng), paired gRNA plasmid, paired gRNA amplicon; 375, open triangle, open circle; 750, black triangle, black circle; 1,500, grey triangle, grey circle; 3,000, black triangle with white line, black circle with white line; respectively. The data in FIG. 42B demonstrate comparable if not higher editing efficiencies for the paired gRNA expression amplicons versus the paired gRNA expression plasmid.

[00265] For the PAM screen (FIG. 37D, FIG. 37E, FIG. 37F, FIG. 39A to FIG. 39D, FIG. 40C. and FIG. 40F), typically, cells were nucleofected with 3 μg of FokI-Cascade polycistronic plasmid and either 150 ng (FokI-PseCascade and FokI-EcoCascade) or ~ 80-120ng (FokI-SthCascade) of oligo-templated paired gRNA expression amplicon (unless otherwise indicated).

[00266] For the specificity analysis (FIG. 38A to FIG. 38C), cells were nucleofected with 3 μg of polycistronic Cascade and 150 ng of oligo-templated paired gRNA expression amplicon and harvested 5 days after nucleofection. At the top of FIG. 38A, the horizontal line represents the interspacer distance, the scissors indicate the expected cut site and the half-sites of the genomic target are with their corresponding PAM regions are shown (Fig. 38A, rectangular boxes with contrasting ends). The relationships of the illustrated half-sites to the target are shown by dashed lines. For each target 32 base pairs are illustrated and the PAM region is shown adjacent the seed sequence. FIG. 38A provides paired gRNAs designed to contain mismatches to one or both half-sites within a genomic target, as denoted by the filled boxes (excluding the PAM sites) in the grids. Note that both half-sites are displayed in the same directionality for simplicity. FIG. 38B provides relative editing efficiency at genomic target 70 for each combination of mismatched paired gRNAs plotted as a percentage of the editing efficiency for the perfectly matching gRNAs. In FIG. 38B, the top line indicates the target (FIG. 38B, Target 70), the next line represents the guide (FIG. 38B, gRNA1 and gRNA2), the next line identifies the mismatched set (FIG. 38B, mm set 1 and mm set 2), the next line illustrates the FokI-Cascade RNP complexes. The left column presents data for relative editing guide 1-mm set1/guide 2-mm set 2, the right column for guide 1-mm set 2/guide 2-mm set 1,

both columns of data present the relative editing efficiency percent (FIG. 38B Relative editing eff (%); scale 0-100), that is, the left column show data for $gRNA_1$ and $gRNA_2$ with mismatched (mm) sets 1 and 2, and the right column shows data for the same target but with swapped mismatched (mm) sets between $gRNA_1$ and $gRNA_2$ (n=1). FIG. 38C provides editing efficiency at target 73 (n=1), displayed as in FIG. 38B..

[00267] After developing the scalable method of generating paired gRNA expression cassettes by oligo-templated PCR amplification (as described herein), which eliminated the need for labor-intensive cloning steps, FokI linker and DNA interspacer lengths were rescreened across a panel of 96 genomic target sites for each homolog variant. With the 17-aa linker, FokI-PseCascade consistently yielded, on average, ~15-25% editing efficiencies within an approximately 30–33 bp interspacer window, and some targets exhibited up to ~40-50% indels (FIG. 37C). Similar trends were observed with the other homologs. PAM requirements were investigated by targeting genomic sites that harbored one cognate PAM and a second mutated PAM. PAM recognition had been shown *in vitro* to be far more promiscuous than the rigid 5'-GG-3' *Streptococcus pyogenes* (*S. pyogenes*) PAM requirement (see, e.g., Szczelkun, M., et al., Proc. Natl. Acad. Sci. USA 111:9798–9803 (2014); Hayes, R., et al., Nature 530:499–503 (2016); Westra, E., et al., Mol. Cell. 46:595–605 (2012); Fineran, P., et al., Proc. Natl. Acad. Sci. USA 111:E1629–E1638 (2014); Leenay, R., et al., Mol. Cell. 62:137–147 (2016)). Strikingly *in vitro* data demonstrated a large number of PAMs were indeed permissive for activity, with a clear rank-order preference emerging (FIG. 37D; FIG. 39A to FIG. 39D). In contrast, editing was completely abolished when the mutated PAM represented a "'self"' target from the CRISPR array.

[00268] In each of FIG. 39A to FIG. 39D, the vertical axis corresponds to editing efficiency (Editing Efficiency (%)) and the horizontal axis corresponds to the PAM sequence associated with the target. FIG. 39A provides FokI-PseCascade editing efficiency as a function of PAM sequence. Genomic sites contained one fixed ATG PAM and a variable PAM at the second half-site, as shown on the horizontal axis. Bars show the mean and s.d. (6-14 sites per variable PAM, n=1 per target site). Note that FIG. 37D describes data for FokI-PseCascade where one PAM is fixed at AAG and the other PAM is variable across a set of PAMs, including ATG. Thus, a subset of those PAMs are AAG-ATG. FIG. 39A describes data for FokI-PseCascade where one PAM is fixed at ATG and the other PAM is variable (FIG. 39A, horizontal axis, left to right, AAG, AAC, AAA, ATG, GAG, ATA, AAT, and AGG) across the set of PAMs, including AAG. Thus, a subset of those PAMs are also AAG-ATG and are the same AAG-ATG sites in FIG. 37D.

[00269] FIG. 39B provides FokI-EcoCascade editing as a function of PAM sequence (FIG. 39B, horizontal axis, left to right, CCG, CGC, AAG, AGG, ATG, GAG, AAA, AAC, ATA, and AAT). The fixed PAM was AAG and bars show the mean and s.d. (6-15 sites per variable PAM, n=1 per target site). FIG. 39C (FIG. 39C, horizontal axis, left to right, AAG, ATG, AAC, AAA, AGG, GAG, AAT, and ATA) provides a similar analysis to that shown FIG. 39B, but the first PAM was fixed to ATG (6-14 sites per variable PAM, n=1 per target site. The ATG column in FIG. 39B, corresponding to an AAG-ATG pair (mean of ~3) is identical to the AAG column in FIG. 39C, corresponding to an AAG-ATG pair (mean also of ~3). Note that the vertical-axes are of different scale. FIG. 39D provides FokI-SthCascade editing as a function of PAM sequence (FIG. 39D, horizontal axis, left to right, CC, AA, GA, TA, and CA). The fixed PAM was GAA, and bars show the mean and s.d. (18-33 sites per variable PAM; n=1 per target site).

[00270] FIG. 40A, FIG. 40B, FIG. 40C, FIG. 40D, FIG. 40E, and FIG. 40F present data related to exemplary changes in editing efficiency of engineered Type I CRISPR-Cas complexes. The data presented in FIG. 40A (FokI-PseCascade) and FIG. 40D (FokI-SthCascade) for percentage editing efficiency (vertical axis) versus interspacer distance in bps (horizontal axis) was obtained essentially as described in Example 20C for the data presented in FIG. 41A and FIG. 41C. In FIG. 40A and FIG. 40D, the horizontal axis represents 23-34 bp interspacer distances, and the bars of the graph, left to right, are FokI-Cas8 polypeptide linker lengths of 17 aa (light grey bars), 20 amino acids (dark grey bars) and 30 aa (white bars). The data presented in FIG. 40C and FIG. 40F was obtained essentially as described for FIG. 39B. FIG. 40C and FIG. 40F provide FokI-PseCascade and FokI-SthCascade editing (FIG. 40C, FIG. 40F, vertical axis, Editing Efficiency (%)) as a function of PAM sequences (FIG. 40C, left to right, CCG, CGC, AAG, AAA, ATG, AAC, AGG, ATA, GAG, and AAT; FIG. 40F, left to right, CC, AA, GA, TA, and CA). FIG. 40B illustrates the FokI-PseCascade RNP complexes. The fixed PAM for FokI-PseCascade was AAG (FIG. 40B, AAG PAM) and the other PAM is variable across a set of PAMs (FIG. 40B, variable PAM). FIG. 40E illustrates the FokI-SthCascade RNP complexes. The fixed PAM for FokI-SthCascade was GAA (FIG. 40B, GAA PAM) and the other PAM is variable across a set of PAMs (FIG. 40E, variable PAM). FokI-PseCascade was rescreened for linker and interspacer preference, and the data demonstrated nearly 50% editing. PAM preference was also examined. From this data, an *in vitro* rank order preference of PAMs was determined. Essentially the same analysis was performed for a variant from *Streptococcus thermophilus*. Editing was lower in the *S. thermophilus* system. However, the data presented herein demonstrates that *in vivo*, in human cells, the PAM preference for the

S. thermophilus system is very promiscuous. The fact that a single A upstream of the protospacer (i.e., target sequence) was permissive for editing, generally provides an increased number of potential target sequences within a gene (e.g., relative to the number of potential Class 2 Type II CRISPR-Cas9 PAM-associated target sites within the same gene). Furthermore, the *in vivo* data presented herein correlates with the *in vitro* PAM preferences demonstrated by Sinkunas, T., et al., EMBO J. 32:385-394 (2013).

[00271] The accumulation of NGS data across hundreds of edited genomic sites provided the ability to characterize DNA repair outcomes of DSBs introduced by FokI-PseCascade. Focusing on 40 unique sites with indel frequencies >10%, the frequencies of deletions and insertions were analyzed as a function of total mutant reads within a 50 bp window surrounding a predicted cleavage site. Insertions of 2–4 bp were highly enriched and present in the vast majority of sites examined (FIG. 37E). Detailed inspection showed that ~90% of these insertions contained perfect duplications of sequences adjacent to the cleavage site. Although not wishing to be limited by any particular theory, such duplications may be the consequence of templated repair of staggered cuts introduced by dimeric FokI.

[00272] The specificity of FokI-PseCascade was evaluated by editing two high-efficiency target sites with an extensive panel of mismatched paired gRNAs (FIG. 38A). Previous studies of Cascade have highlighted an ~8-nt PAM-proximal seed sequence, as well as mismatch promiscuity at every 6th position within the 32-nt guide gRNA, due to these bases being flipped out of the RNA-DNA heteroduplex structure formed upon target binding (see, e.g., Jung, C., et al., Cell 170:35–47 (2017); Mulepati, S., et al., Science 345:1479–1484 (2014); Fineran, P., et al., Proc. Natl. Acad. Sci. USA 111:E1629–E1638 (2014); Semenova, E., et al., Proc. Natl. Acad. Sci. USA 108:10098–10103 (2011)). Mismatches within the PAM-proximal seed region were highly deleterious for genome editing, whereas mismatches distal from the PAM were well tolerated, leading to near-wild-type editing efficiencies (FIG. 38B; FIG. 38C). When blocks of mismatches were present in both half-sites, however, editing dropped dramatically across the entire panel of paired gRNAs tested (FIG. 38B, FIG. 38C). Based on the data on the PAM and interspacer data of FokI-PseCascade-mediated genome editing (FIG.38C; FIG. 37D), one advantage of the engineered Type I CRISPR-Cas complexes of the present invention is that a targetable site can occur every ~20 to ~30 bp in the human genome, whereas editing at potential off-target sites is unlikely.

[00273] Accordingly, in one embodiment of the present invention, the potential targetable sites, or "target density," of a given engineered FokI-Cascade system is a function of its efficient interspace distance and PAM preference, and will have some variability across

homologs. In some embodiments, the following criteria can be used to calculate the target density in the human genome for FokI-PseCascade, FokI-EcoCascade, and FokI-SthCascade (the data were extrapolated to calculate predicted target density).

[00274] FokI-PseCascade, target density can be calculated using the following motif:

[00275] 5'–[half-site$_1$–PAM$_1$]–[interspacer]–[PAM$_2$–half-site$_2$]–3'.

[00276] Here, [half-site$_1$–PAM$_1$] denotes the reverse-complement of the half-site$_1$ gRNA$_1$ target-strand target sequence and PAM, and [half-site$_2$–PAM$_2$] denotes the half-site$_2$ gRNA$_2$ non-target strand PAM and target-sequence. Based on the distribution of interspacer lengths that supported editing with FokI-PseCascade (see, e.g., FIG. 37D), an efficient interspacer length is about 30-33 bp. PAMs were defined as belonging to either set 1 which gave the highest editing (AAG, AAA, ATG, AAC) or to set 2 if they contained any of the tested PAMs that showed activity (AAG, AGG, ATG, GAG, AAA, AAC, AAT, ATA) (see, e.g., FIG. 39A; FIG, 40B). From this, potential target sites satisfying the preferred interspacer length criterion with two PAMs belonging to either set 1 or set 2 will occur on average every 33.4 bp or 9.2 bp, respectively.

[00277] The target density for FokI-EcoCascade was determined similarly, except the interspacer length was defined as 31-33 and PAMs were defined as belonging to either set 1, which gave the highest editing (AAG, AGG, ATG, GAG, AAA), or set 2 if they contained any of the tested PAMs that showed activity (AAG, AGG, ATG, GAG, AAA, AAC, AAT, ATA) (see, e.g., FIG. 39C; FIG. 39D). From this, potential target sites were calculated with set 1 PAMs or set 2 PAMs occurring, on average, every 30.4 bp or 12.2 bp, respectively.

[00278] The human genome target density for FokI-SthCascade was determined similarly, except the interspacer length was defined as 29-31 bp and PAMs were defined as NNA (see, e.g., FIG. 39D). From this, potential target sites were calculated to occur, on average, every 4 bp.

[00279] Accordingly, engineered Type I CRISPR-Cas complexes, as described herein, provide a method to provide a variety of potential target sites by providing a number of PAM-adjacent target sequences available for genomic editing. Thus, one embodiment of the present invention relates to a method of using PAM sequences associated with engineered Type I CRISPR-Cas complexes to provide an increased number of available target sequences within a gene (e.g., relative to the number of available target sequences associated with PAM sequences of Class 2 CRISPR-Cas Type II or Type V systems). Applications of this method relate to use of engineered Type I CRISPR-Cas complexes that can include, but are not limited to, binding to and/or cleavage of a target sequence, mutation of a target sequence, transcriptional regulation

related to a target sequence or regulatory elements thereof, as well as intentional modification, change, and/or markedly different structural change (e.g., in a product of the gene) mediated by use of the engineered Type I CRISPR-Cas complexes described herein.

[00280] In some embodiments, the engineered Type I CRISPR-Cas effector complexes described herein can be used to generate non-human transgenic organisms by site specifically introducing a selected polynucleotide sequence (e.g., a portion of a donor polynucleotide) at a DNA target locus in the genome to generate a modification, change, and or mutation of the gDNA. The transgenic organism can be an animal or a plant.

[00281] A transgenic animal is typically generated by introducing engineered Type I CRISPR-Cas effector complexes into a zygote cell. A basic technique, described with reference to making transgenic mice (see, e.g., Cho, A., et al., "Generation of Transgenic Mice," Current Protocols in Cell Biology, CHAPTER.Unit-19.11 (2009)) involves five basic steps: first, preparation of a system, as described herein, including a suitable donor polynucleotide; second, harvesting of donor zygotes; third, microinjection of the system into the mouse zygote; fourth, implantation of microinjected zygotes into pseudo-pregnant recipient mice; and fifth, performing genotyping and analysis of the modification of the gDNA established in founder mice. The founder mice will pass the genetic modification to any progeny. The founder mice are typically heterozygous for the transgene. Mating between these mice will produce mice that are homozygous for the transgene 25% of the time.

[00282] Methods for generating transgenic plants are also well known and can be applied using engineered 1 Type I CRISPR-Cas effector complexes. A generated transgenic plant, for example using *Agrobacterium*-mediated transformation, typically contains one transgene inserted into one chromosome. It is possible to produce a transgenic plant that is homozygous with respect to a transgene by sexually mating (i.e., selfing) an independent segregant transgenic plant containing a single transgene to itself. Typical zygosity assays include, but are not limited to, single nucleotide polymorphism assays and thermal amplification assays that distinguish between homozygotes and heterozygotes.

[00283] In a sixth aspect, the present invention relates to use of engineered Type I CRISPR-Cas effector complexes to create substrate channels. In some embodiments, fusion proteins comprising substrate channel elements and Cas7 subunit proteins are constructed. These Cas7 fusion proteins are then assembled into an engineered Type I CRISPR-Cas effector complex (e.g., comprising Cse2, Cas5, Cas6, Cas7-substrate channel element fusions, and Cas8). In some embodiments, the crRNA of the engineered Type I CRISPR-Cas effector complex can be extended to accommodate additional Cas7 subunits (see, e.g., Luo, M., et al.,

Nucleic Acids Res. 44:7385-7394 (2016)). Different substrate elements can be fused to Cas7 and then mixed at the desired stoichiometry. When these various Cas7 subunits assemble into a complete Type I CRISPR-Cas effector complex, co-localization of substrate elements can enhance the efficacy of substrate channeling.

[00284] In some embodiments, an RNA scaffold is constructed such that multiple Cas7-substrate channel element fusions can bind to it in the absence of other Type I CRISPR-Cas effector complex components.

[00285] Substrate channel elements can be fused to the N-terminus of Cas7 and/or the C-terminus of Cas7. In addition, circular permutations of Cas7 can be fused to substrate channel elements.

[00286] FIG. 11A and FIG. 11B presents illustrations of substrate channels consisting of three consecutive enzymes in a pathway. Substrate channels facilitate the passing of intermediary metabolic products directly to the active site of the consecutive enzyme in the metabolic pathway chain without release into the extra channel space. FIG. 11A illustrates a typical arrangement of an engineered substrate channel. Enzymes E1, E2, and E3 interact covalently or non-covalently to a scaffold protein (S1, S2, S3) matrix. The double-headed arrows represent interactions (e.g., affinity interactions) between an enzyme and a scaffold protein. The substrate (X) is then processed to the product (Y) without release to the extra channel space. FIG. 11B illustrates one embodiment of the present invention comprising an engineered Type I CRISPR-Cas effector complex that carries enzymes E1, E2, and E3 as fusion proteins to Cas7 subunit proteins (i.e., a covalent interaction), thus creating a substrate channel. cpCas7 proteins and backbones formed of cpCas7 proteins can also be useful in the practice of this aspect of the present invention.

[00287] In other embodiments, substrate channel elements can be fused to Cas6. The Cas6 subunit of Cascade complexes recognizes specific RNA hairpin structures. An RNA scaffold can be constructed that is composed of multiple Cas6 RNA hairpin structures concatenated together. Cas6 peptides from different Cascade complexes have different recognition sequences. Accordingly, RNA scaffolds can be constructed from multiple orthogonal Cas6 RNA hairpins. By fusing different substrate channel elements to orthogonal Cas6 peptides, substrate channel complexes can be assembled in specific stoichiometry.

[00288] Substrate channel elements can be fused to the N-terminus of Cas6 and/or the C-terminus of Cas6. In addition, circular permutations of Cas6 can be fused to substrate channel elements.

[00289] In some embodiments, a heterologous metabolic pathway of interest can be expressed in a model organism, such as *E. coli*. When genes are heterologously expressed, the genes can be codon-optimized to express the genes more efficiently.

[00290] In one embodiment, the metabolic pathway of interest is the mevalonate pathway from *Saccharomyces cerevisiae*. Substrate channel elements of this pathway include, but are not limited to, acetoacetyl-CoA-thioase (AtoB), hydroxy-methylglutaryl-CoA synthase (HMGS), and hydroxy-methylglutaryl-CoA reductase (HMGR).

[00291] In another embodiment, the metabolic pathway of interest is the glycerol synthesis pathway from *S. cerevisiae*. Substrate channel elements of this pathway include, but are not limited to, glycerol-3-phosphate dehydrogenase (GPD1) and glycerol-3-phosphate phosphatase (GPP2).

[00292] In yet another embodiment, the metabolic pathway of interest is the starch hydrolysis pathway from *Clostridium stercorarium*. Substrate channel elements of this pathway include, but are not limited to, CelY and CelZ.

[00293] In an additional embodiment, the metabolic pathway of interest is the glucose phosphotransferase pathway from *E. coli*. Substrate channel elements of this pathway include, but are not limited to, trehalose-6-phosphate synthetase (TPS) and trehalose-6-phosphate phosphatase (TPP).

[00294] In a seventh aspect, the present invention relates to site-directed recruitment of functional domains fused to Cascade subunit proteins by complexes comprising a Class 2 Type II Cas9 protein and a nucleic acid-targeting nucleic acid (NATNA). Functional domains are disclosed herein and include, but are not limited to, protein domains having enzymatic function, capable of transcriptional activation, or capable of transcriptional repression. Example 13A and Example 13B describe a method of engineering a Class 2 Type II CRISPR sgRNA, crRNA, tracrRNA, or crRNA and tracrRNA sequences with a Class 1 Type I CRISPR repeat stem sequence, allowing for the recruitment of one or more Cascade subunit proteins to a Type II CRISPR Cas protein/guide RNA complex binding site.

[00295] FIG. 12A, FIG. 12B, and FIG. 12C present a generalized illustration of the site-directed recruitment of a functional protein domain fused to a Cascade subunit protein by a dCas9:NATNA complex to a target site. A Class 2 Type II CRISPR NATNA (FIG. 12A, 102) comprising a spacer sequence (FIG. 12A, 101) is covalently linked through a linker nucleic acid sequence (FIG. 12A, 103) to a Class 1 Type I CRISPR repeat stem sequence (FIG. 12A, 104). The Type II CRISRP NATNA covalently linked to the Type I CRISPR repeat stem sequence (FIG. 12A, 105) is capable of binding to a Type II dCas9 (FIG. 12A, 106) and a Type I Cascade

subunit protein (e.g., Cas6; FIG. 12A, 107), which is fused though a linker sequence (FIG. 12A, 108) to a functional protein domain (e.g., an enzymatic domain, a transcriptional activation or repression domain; FIG. 12A, 109) to form an RNP complex. This RNP complex (FIG. 12B, 110) is capable of targeting a double-stranded DNA (FIG. 12B, 111) comprising a target sequence (FIG. 12B, 112) complementary to the Type II CRISPR NATNA spacer sequence (FIG. 12A, 101). Target recognition by the RNP complex results in hybridization (FIG. 12B, 113) between the spacer sequence (FIG. 12A, 101) and the target sequence (FIG. 12B, 112). Localization of the Cascade subunit-functional domain fusion protein to the DNA allows for modification of the DNA by the functional protein domain or transcriptional regulation of an adjacent gene (FIG. 12C, 114).

[00296] In an eighth aspect, the present invention relates to compositions comprising engineered Type I CRISPR-Cas effector complexes, engineered guide polynucleotides, and combinations thereof. In some embodiments, the engineered Type I CRISPR-Cas effector complex comprises an associated Cas3 fusion protein. Wild-type Type I CRISPR-Cas systems require coordinated action of the Cascade effector complex for DNA targeting and the Cas3 helicase-nuclease for processive DNA degradation. In one embodiment of the present invention, Type I CRISPR-Cas effector complexes were engineered to make precise DSBs by fusing the complex to a nuclease domain (e.g., a non-specific FokI endonuclease domain). This approach uses paired guide polynucleotides that target two half-site DNA sequences separated by an intervening sequence (i.e., the interspacer).

[00297] An embodiment of this aspect of the present invention relates to a composition comprising two engineered Type I CRISPR-Cas effector complexes each comprising a spacer and a fusion protein comprising a Cas subunit and an endonuclease (e.g., a FokI; see, e.g., the Cascade complexes of FIG. 2A, FIG. 2B, and FIG. 2C), wherein at least two parameters are varied to modulate genome editing efficiency. Such parameters include:

[00298] the length of a linker polypeptide used to produce the fusion protein comprising a Cas subunit protein and the endonuclease (e.g., FokI); and

[00299] the length of the interspacer distance between the nucleic acid target sequences to which the spacers are capable of binding.

[00300] Guidance is provided herein regarding the amino acid composition and sequence linker polypeptides.

[00301] One embodiment of this aspect of the present invention is a composition comprising:

[00302] a first engineered Type I CRISPR-Cas effector complex comprising,

[00303] a first Cse2 subunit protein, a first Cas5 subunit protein, a first Cas6 subunit protein, and a first Cas7 subunit protein,

[00304] a first fusion protein comprising a first Cas8 subunit protein and a first FokI, wherein the N-terminus of the first Cas8 subunit protein or the C-terminus of the first Cas8 subunit protein is covalently connected by a first linker polypeptide to the C-terminus or N-terminus, respectively, of the first FokI, and wherein the first linker polypeptide has a length of between about 10 amino acids and about 40 amino acids, and

[00305] a first guide polynucleotide comprising a first spacer capable of binding a first nucleic acid target sequence; and

[00306] a second engineered Type I CRISPR-Cas effector complex comprising,

[00307] a second Cse2 subunit protein, a second Cas5 subunit protein, a second Cas6 subunit protein, and a second Cas7 subunit protein,

[00308] a second fusion protein comprising a second Cas8 subunit protein and a second FokI, wherein the N-terminus of the second Cas8 subunit protein or the C-terminus of the second Cas8 protein is covalently connected by a second linker polypeptide to the C-terminus or N-terminus, respectively, of the second FokI, and wherein the second linker polypeptide has a length of between about 10 amino acids and about 40 amino acids, and

[00309] a second guide polynucleotide comprising a second spacer capable of binding a second nucleic acid target sequence, wherein a protospacer adjacent motif (PAM) of the second nucleic acid target sequence and a PAM of the first nucleic acid target sequence have an interspacer distance between about 20 base pairs and about 42 base pairs.

[00310] Examples of such a first engineered Type I CRISPR-Cas effector complex bound to a first nucleic acid target sequence and a second engineered Type I CRISPR-Cas effector complex bound to a second nucleic acid target sequence are illustrated in FIG. 2A, FIG. 2B, and FIG. 2C.

[00311] In some embodiments, the length of the first linker polypeptide and/or the second linker polypeptide is a length of between about 15 amino acids and about 30 amino acids, or between about 17 amino acids and about 20 amino acids. In one embodiment, the length of the first linker polypeptide and the second linker polypeptide are the same.

[00312] The first Cas8 subunit protein and the second Cas8 subunit protein can each comprise identical amino acid sequences of the Cas8 subunit protein.

[00313] Similarly, the first Cse2 subunit protein and the second Cse2 subunit protein can each comprise identical amino acid sequences of the Cse2 subunit protein, the first Cas5 subunit protein and the second Cas5 subunit protein can each comprise identical amino acid

sequences of the Cas5 subunit protein, the first Cas6 subunit protein and the second Cas6 subunit protein can each comprise identical amino acid sequences of the Cas6 subunit protein, the first Cas7 subunit protein and the second Cas7 subunit protein can each comprise identical amino acid sequences of the Cas7 subunit protein, and combinations thereof.

[00314] Typically, the N-terminus of the first Cas8 subunit protein is covalently connected by the first linker polypeptide to the C-terminus of the first FokI, the C-terminus of the first Cas8 subunit protein is covalently connected by a first linker polypeptide to the N-terminus of the first FokI, the N-terminus of the second Cas8 subunit protein is covalently connected by the second linker polypeptide to the C-terminus of the second FokI, the C-terminus of the second Cas8 subunit protein is covalently connected by a second linker polypeptide to the N-terminus of the second FokI, and combinations thereof.

[00315] Embodiments of this aspect of the present invention include embodiments wherein the length between the second nucleic acid target sequence and the first nucleic acid target sequence is an interspacer distance between about 22 base pairs and about 40 base pairs, between about 26 base pairs and about 36 base pairs, between about 29 base pairs and about 35 base pairs, or between about 30 base pairs and about 34 base pairs.

[00316] The first FokI and the second FokI can be monomeric subunits that are capable of associating to form a homodimer, or distinct subunits that are capable of associating to form a heterodimer.

[00317] In a preferred embodiment, the guide polynucleotides comprise RNA.

[00318] In some embodiments, gDNA comprises the PAM of the second nucleic acid target sequence and the PAM of the first nucleic acid target sequence.

[00319] In some embodiments, the engineered Type I CRISPR-Cas effector complexes are based on Type I CRISPR-Cas effector complexes of one or more organisms selected from the group consisting of *Salmonella enterica*, *Geothermobacter sp.* (strain EPR-M), *Methanocella arvoryzae MRE50*, *Streptococcus thermophilus* (e.g., *Streptococcus thermophilus* (strain ND07), *Pseudomonas sp. S-6-2*, and *E. coli*. In preferred embodiments, the engineered Type I CRISPR-Cas effector complexes are based on Type I CRISPR-Cas effector complexes of *Streptococcus thermophilus* (e.g., *Streptococcus thermophilus* (strain ND07), *Pseudomonas sp. S-6-2,* and/or *E. coli. Pseudomonas sp.* S-6-2 induced ~10-fold higher editing efficiencies than the *E. coli* homolog, and roughly one half of the other homologs tested showed activities on par with *E. coli*, demonstrating that engineered Type I CRISPR-Cas effector complexes from diverse Type I systems can be functionally used for genome editing in human cells.

[00320] The data presented in Example 18A, Example 18B, Example 18C, Example 18D, Example 20A, Example 20B, and Example 20C demonstrate that varying the length of the linker polypeptide used to produce the fusion protein comprising the Cas subunit protein and the FokI and/or varying the length of the interspacer distance between the nucleic acid target sequences to which the spacers are capable of binding facilitate modulation of genome editing efficiency in cells.

[00321] In yet another embodiment, the present invention relates to an engineered Type I CRISPR-Cas effector complex comprising a first fusion protein that comprises a Cascade subunit protein (e.g., a Cas8 subunit protein) and a first functional domain (e.g., FokI), and a second fusion protein that comprises a dCas3* protein and a second functional domain (e.g., FokI) (FIG.13A: Cas7, Cas5, Cas8, Cse2, and Cas6, the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin). The engineered Type I CRISPR-Cas effector complex comprising the first functional domain (e.g., FokI) (FIG. 13A, Cas8-linker1-FP1 fusion) can bind DNA and can then recruit the dCas3*-second functional domain (e.g., FokI) fusion protein (FIG. 13A, dCas3*-linker2-FP2). In the case where the first functional domain (FIG. 13A, Cas8-linker1-FP1 fusion) and the second functional domain (FIG. 13A, dCas3*-linker2-FP2) comprise subunits of a dimeric protein, the dCas3*-second functional domain (e.g., FokI) fusion protein binds the engineered Type I CRISPR-Cas effector complex comprising the first functional domain (e.g., FokI) facilitating dimerization of the first functional domain and the second functional domain (FIG. 13A). FIG. 14A illustrates the binding to dsDNA of an engineered Type I CRISPR-Cas effector complex (FIG. 14A, Cascade) comprising the first functional domain (FIG. 14A, FD1) connected to a Cas subunit protein (FIG. 14A, striped box) via a linker polypeptide (FIG. 14A, Linker 1) and a dCas3* connected to a second functional domain (FIG. 14A, FD2) via a linker polypeptide (FIG. 14A, Linker 2) associated with the Cascade complex; thus bringing FD1 and FD2 into proximity and facilitating the interaction of FD1 and FD2. Binding of the Cascade complex involves a single PAM sequence (FIG. 14A, PAM, open box). In FIG. 14A, dsDNA is illustrated as paired, horizontal dashed lines. In the case of the functional domain being a dimeric endonuclease (e.g., FokI), the proximity of FD1 and FD2 facilitates formation of a functional dimer.

[00322] One advantage of this embodiment of the present invention is that a single Cascade complex (recognizing a single PAM sequence) can be used to cleave a double-stranded nucleic acid target sequence, versus using two FokI-Cascade complexes (compare FIG. 14A with FIG. 2A, FIG. 2B, and FIG. 2C). Using two FokI-Cascade complexes requires two PAM

sequences in the proper orientation (FIG. 2A, FIG. 2B, and FIG. 2C), which can limit selection of proximal nucleic acid target sequences.

[00323] The length and/or composition of the linker polypeptide used to produce the fusion protein comprising a Cas subunit protein and an endonuclease (e.g., FokI), as well as the length and/or composition of the linker polypeptide used to produce the fusion protein comprising a dCas3* protein and an endonuclease, can be varied to modulate genome editing efficiency. Example 21A, Example 21B, Example 21C, and Example 21D describes the design and testing of multiple Cas3-FokI linker compositions and lengths and FokI-Cas8 linker compositions and lengths for modulation of genome editing efficiency.

[00324] Another embodiment of this aspect of the invention comprises an engineered Type I CRISPR-Cas effector complex (FIG. 13B: Cas7, Cas5, Cas8, Cse2, and Cas6; the dashed box around Cas6 indicates its interaction with the crRNA hairpin; the cRNA is illustrated as a black line comprising the hairpin) and a fusion protein comprising a dCas3* protein (FIG. 13B, dCas3*) and a functional domain (FIG. 13B, FP) (e.g., cytidine deaminase) connected by a linker polypeptide (FIG. 13B, Linker). The engineered Type I CRISPR-Cas effector complex can bind DNA and recruit the dCas3*-functional domain (e.g., cytidine deaminase) fusion protein. This embodiment can facilitate site-specific targeting of a nucleic acid target sequence for modification by, or interaction with, a functional domain. In the case of cytidine deaminase, an engineered Type I CRISPR-Cas effector complex and a fusion protein that comprises a dCas3* protein and cytidine deaminase can be used for site-specific base editing in a nucleic acid target sequence. FIG. 14B illustrates an example of an engineered Type I CRISPR-Cas effector complex (FIG. 14B, Cascade) comprising a fusion protein comprising a dCas3* protein (FIG. 14B, dCas3*) connected with a functional domain (FIG. 14B, FD) via a linker polypeptide (FIG. 14B, Linker), wherein the complex is bound to dsDNA (FIG. 14B, paired, horizontal dashed lines). In FIG. 14B, contact of the functional domain with dsDNA is facilitated. Binding of the Cascade complex involves a single PAM sequence (FIG. 14B, PAM, open box). FIG. 14C illustrates another example of an engineered Type I CRISPR-Cas effector complex (FIG. 14C, Cascade) comprising a fusion protein comprising a dCas3* protein (FIG. 14C, dCas3*) connected with a functional domain (FIG. 14C, FD) via a linker polypeptide (FIG. 14C, Linker), wherein the complex is bound to dsDNA (FIG. 14C, paired, horizontal dashed lines). Binding of the Cascade complex involves a single PAM sequence (FIG. 14C, PAM, open box). In FIG. 14C, contact of the functional domain with ssDNA is facilitated.

[00325] Additional functional domains and proteins that can be used to construct fusion proteins with Type I CRISPR-Cas subunit proteins are described in the present Specification

and Examples. Linker polypeptide compositions and lengths for Cas3-linker polypeptide-functional domain fusion proteins can be evaluated following the guidance of Example 21A to Example 21D and the present Specification to evaluate effects on the performance of the functional domain.

[00326] Some embodiments of the present invention can use an engineered Type I CRISPR-Cas effector complex and a mCas3 protein, wherein the mCas3 protein comprises down-modulated helicase activity (e.g., the mCas3 protein, a Cas3 processivity mutant protein, has reduced movement along DNA relative to a wild-type Type I CRISPR Cas3 protein) or the mCas3 protein lacks helicase activity (e.g., the mCas3 protein is no longer a processive nuclease like wtCas3 protein, but the mCas3 protein retains nicking activity). The engineered Type I CRISPR-Cas effector complexes can bind DNA and then recruit the mCas3 protein. This embodiment can facilitate site-specific cleavage of genomic DNA.

[00327] Table 48 describes a number of mCas3 proteins, wherein the mutations made to the Cas3 protein affected the ATP binding/hydrolysis region of the helicase domain or the ssDNA path conserved region of the helicase domain. FIG. 44 shows a linear representation of the functional domains of the EcoCas3 protein and the relative locations of mutants made within the Cas3 coding sequence. In FIG. 44, the HD nuclease domain (amino acids 1-272), Helicase domain (RecA1 region, amino acids 273-521; RecA2 region, amino acids 522-737), Linker (amino acids 738-794), and C-terminal domain (CTD, amino acids 795-888) are indicated. Huo, Y., et. al., Nat. Struct. Mol. Biol. 9:771-777 (2014) provide a sequence conservation analysis with sequence alignments of the Cas3 family of proteins from *Thermobifida fusca* (accession code: Q47PJ0; SEQ ID NO:1869), *Saccharomonospora viridis* (C7MTA6; SEQ ID NO:1870), *Thermomonospora curvata* (D1A6Q2; SEQ ID NO:1922), *Streptomyces avermitilis* (Q825B5; SEQ ID NO:1925), *Streptomyces bottropensis* (M3DI13; SEQ ID NO:1923), *Thermus thermophilus strain HD8* (Q53VY2; SEQ ID NO:1924) and *E. coli* (P38036; SEQ ID NO:1844). 24 different EcoCas3 protein variants with mutations in the ATP binding portion of the helicase domain or ssDNA loop binding domain were screened (Example 23A to Example 23C). Several mutants showed significantly more and/or position-shifted deletion classes within the amplicon window; a finding which supports that those mCas3 proteins had reduced processivity relative to wtCas3.

[00328] Example 23A to Example 23C describe such mCas3 proteins, wherein the average mCas3 protein-induced deletions are shorter relative to the average deletions generated with the corresponding wtCas3 protein. Such mCas3 proteins are useful for genome editing (e.g., in human cells). FIG. 45A, FIG. 45B, FIG. 45C, and FIG. 45D present data indicative of

mCas3 proteins that, when associated with Cascade RNP complexes, generate shorter average deletion lengths relative to wtCas3 protein, in association with a Cascade RNP complex, when introduced into and expressed in human cells. In view of the teachings of the present Specification, one of ordinary skill in the art can make similar mutations in the corresponding regions of Cas3 proteins obtained from other species of bacteria in addition to *E. coli*.

[00329] Example 26A to Example 26C provide an additional example of a mCas3 protein useful for generating genomic deletions, wherein the average mCas3 protein-induced deletions are shorter relative to the average deletions generated with the corresponding wtCas3 protein. The data presented in the example support that an ATPase/helicase deficient variant of Cas3 from *Pseudomonas* sp. S-6-2 (mPseCas3 protein) can be used with PseCascade RNP complexes to generate deletions at the expected cleavage site (i.e., cleavage site localized deletion).

[00330] wtPseCas3 protein/PseCascade activity was further characterized. Additional experiments were performed using target-enrichment probes, which enable detection of large genomic deletions. Specifically, HEK293 cells were transfected with DNA templates encoding PseCascade RNP complex, wtPseCas3 protein, and a minimal CRISPR array directed to the TRAC locus essentially as described in Examples 26A to Example 26C. Target-enrichment probes were used to isolate and sequence genomic fragments; whereas in Example 26C, an amplicon window was used to identify the presence of deletions. The target-enrichment/sequencing method provided an unbiased view of larger deletions not provided by using an amplicon window to identify deletions. Overall, deletions evaluated using target-enrichment and sequencing of genomic fragments were found to be largely unidirectional, starting upstream of the wtPseCas3 protein initiation site. The deletions ranged from 1 bp to nearly 250 kb. In addition to providing a method of cutting genomic DNA and providing deletions of a given length, this method may be useful for generating large, random subsets of deletions at defined locations to probe regulatory/promoter regions of genes.

[00331] mCas3 proteins can comprise one or more mutations (e.g., combinations of the mutations as described in Table 48).

[00332] Control of deletion lengths was demonstrated for several mCas3 proteins. In some embodiments, mCas3 proteins of the present invention, in association with a Cascade complex comprising a guide polynucleotide, may provide average deletion lengths of between about 1 and about 600 base pairs, about 1 and about 500 base pairs, about 1 and about 400 base pairs, about 1 and about 300 base pairs, preferably between about 1 and about 250 base pairs, between about 1 and about 200 base pairs, or between about 1 and about 100 base pairs.

[00333] In some embodiments, wtCas3 proteins or mCas3 proteins can be fused to the various subunits of the Cascade complex to further control Cas3 average deletion lengths. Tethering to the Cascade complex may limit or prevent Cas3 protein or mCas3 protein movement along DNA, because as it will be fixed to the locus where the Cascade complex is bound. wtCas3 proteins or mCas3 proteins can be fused, typically with a linker polypeptide, to either the N- or C-terminal domain of protein components of a Cascade complex (e.g., for an EcoCascade complex fusions can be with EcoCas8, EcoCas6, or EcoCas5). NLS sequences can also be appended to the N-terminus of the fusion proteins. Examples of such constructs for *E. coli* Cascade protein components are presented in Table 12. These EcoCas3 fusion proteins also have NLS sequences appended to their N-termini.

[00334]

Table 12
Plasmids Encoding EcoCascade Comprising Cas3 Fusion Proteins

DNA sequences SEQ ID NO:	Corresponding protein sequences* SEQ ID NO:	Cas3 fusion, linker length, and Cascade complex gene	Fusion to N- or C- terminus of Cascade complex gene
1875	1881	Cas3-17aa-Cas8 fusion	N-terminal
1876	1882	Cas8-17aa-Cas3 fusion	C-terminal
1877	1883	Cas3-17aa-Cas5 fusion	N-terminal
1878	1884	Cas5-17aa-Cas3 fusion	C-terminal
1879	1885	Cas3-17aa-Cas6 fusion	N-terminal
1880	1886	Cas6-17aa-Cas3 fusion	C-terminal

* protein sequence is the encoded polycistronic protein sequence

[00335] Embodiments of the present invention include an engineered Type I CRISPR mCas3 protein capable of reduced movement along DNA relative to a wild-type Type I CRISPR Cas3 protein (wtCas3 protein). In some embodiments, the mCas3 protein comprises about 90% or higher, preferably about 95% or higher, more preferably about 98% or higher sequence identity to the corresponding wtCas3 protein. The coding sequence for the mCas3 protein can comprise a nuclear localization signal covalently connected at the amino terminus, carboxy terminus, or both the amino and carboxy termini. A mCas3 protein can comprise one or more mutations that down-modulates helicase activity, wherein the engineered mCas3 protein retains nuclease activity (or at least a portion thereof) relative to the corresponding wtCas3 protein. Typically, DNA is dsDNA comprising a target region comprising a nucleic acid target sequence. When the wtCas3 protein is associated with a corresponding Cascade nucleoprotein complex ("Cascade NP complex/wtCas3 protein"; e.g., a Cascade RNP

complex), and the Cascade NP complex comprises a guide comprising a spacer complementary to the nucleic acid target sequence, binding of the Cascade NP complex/wtCas3 protein to the nucleic acid target sequence facilitates cleavage in the target region of the DNA, typically resulting in a deletion in the target region; and the mCas3 protein when it is associated with the Cascade NP complex ("Cascade NP complex/mCas3 protein"; e.g., a Cascade RNP complex/mCas3 protein) and binds the nucleic acid target sequence facilitates cleavage in the target region of the DNA and results in a shorter average deletion length relative to the wtCas3 average deletion length.

[00336] In some embodiments, the one or more mutations in the mCas3 protein are substitutions of amino acids relative to the wtCas3 protein. In other embodiments, the one or more deletions comprise deletion or insertion of amino acids in the mCas3 protein coding sequence relative to the wtCas3 protein. The one or more mutations can be in either the RecA1 region or RecA2 region of the helicase domain. In one embodiment, the one or more mutations down-modulate binding of the mCas3 protein to ssDNA relative to the wtCas3 protein (e.g., a mutation affecting ssDNA loop binding and/or a mutation in the ssDNA path conserved region of the helicase domain). In additional embodiments, the one or more mutations down-modulate hydrolysis of ATP by the mCas3 protein relative to wtCas3 protein or down-modulate binding of ATP to the mCas3 protein relative to the wtCas3 protein. In a further embodiment, a mCas3 protein comprises combinations of one or more mutations that down-modulate binding of the mCas3 protein to ssDNA relative to the wtCas3 protein, down-modulate hydrolysis of ATP by the mCas3 protein or down-modulate binding of ATP to the mCas3 protein relative to the wtCas3 protein.

[00337] Further embodiments include the coding sequences for the mCas3 protein covalently connected to the amino terminus or carboxy terminus of coding sequences of a Cas protein of the Cascade nucleoprotein complex (e.g, a Cascade RNP complex). Such a Cas protein can be selected from the group consisting of Cse2, Cas8 protein, Cas7 protein, Cas6, and Cas5 protein.

[00338] In some embodiments, the wtCas3 protein is an *E. coli* Type 1 CRISPR Cas3 protein. In other embodiments, the wtCas3 protein is a wtCas3 protein selected from the group consisting of *Pseudomonas sp. S-6-2*, *Thermobifida fusca*, *Saccharomonospora viridis*, *Thermomonospora curvata*, *Streptomyces avermitilis*, *Streptomyces bottropensis*, *Thermus thermophilus*, *Vibrio cholera*, *Salmonella enterica*, *Geothermobacter sp. EPR-M*, *Methanocella arvoryzae MRE50*, and *Streptococcus thermophilus (strain ND07)*.

[00339] For an *E. coli* Type 1 CRISPR wtCas3 protein, the one or more mutations can include, but are not limited to, D452H, A602V, or D452H and A602V.

[00340] In further embodiments, a cell comprises the DNA, wherein the cell can be a eukaryotic cell (e.g., a human cell).

[00341] In additional embodiments, the present invention includes polynucleotides comprising coding sequences for mCas3 proteins, expression cassettes comprising mCas3 protein coding sequences, plasmids comprising mCas3 protein coding sequences, and Cascade nucleoprotein complexes comprising mCas3 proteins.

[00342] In a ninth aspect, the present invention relates to methods of using engineered Type I CRISPR-Cas effector complexes.

[00343] In some embodiments, the present invention includes a method of binding a nucleic acid target sequence in a polynucleotide (e.g., dsDNA) comprising providing one or more engineered Type I CRISPR-Cas effector complexes for introduction into a cell or a biochemical reaction and introducing the engineered Type I CRISPR-Cas effector complexes into the cell or biochemical reaction, thereby facilitating contact of the engineered Type I CRISPR-Cas effector complexes with the polynucleotide. Contact of the complexes with the polynucleotide results in binding of the engineered Type I CRISPR-Cas effector complexes to the nucleic acid target sequence(s) in the polynucleotide.

[00344] In one embodiment, an engineered Type I CRISPR-Cas effector complex comprises a guide complementary to a nucleic acid target sequence in the polynucleotide. The engineered Type I CRISPR-Cas effector complex binds to a nucleic acid target sequence in the polynucleotide.

[00345] In a further embodiment, a first engineered Type I CRISPR-Cas effector complex comprises a guide complementary to a first nucleic acid target sequence in the polynucleotide and a second engineered Type I CRISPR-Cas effector complex comprises a guide complementary to a second nucleic acid target sequence in the polynucleotide. The first engineered 1 Type I CRISPR-Cas effector complex binds to a first nucleic acid target sequence and the second engineered Type I CRISPR-Cas effector complex binds to a second nucleic acid target sequence in the polynucleotide.

[00346] In yet another embodiment, an engineered Type I CRISPR-Cas effector complex comprises a guide complementary to a nucleic acid target sequence in the polynucleotide and further comprises a dCas3* fusion protein capable of associating with the complex. The engineered Type I CRISPR-Cas effector complex binds to a nucleic acid target sequence in the

polynucleotide, and the effector complex comprises a dCas3* fusion protein associated with the complex.

[00347] Such methods of binding a nucleic acid target sequence can be carried out *in vitro* (e.g., in a biochemical reaction or in cultured cells; in some embodiments, the cultured cells are human cultured cells that remain in culture and are not introduced into a human); *in vivo* (e.g., in cells of a living organism, with the proviso that, in some embodiments, the organism is a non-human organism); or *ex vivo* (e.g., cells removed from a subject, with the proviso that, in some embodiments, the subject includes a human subject, and in other embodiments the subject is a non-human subject).

[00348] A variety of methods are known in the art to evaluate and/or quantitate interactions between nucleic acid sequences and polypeptides including, but not limited to, the following: immunoprecipitation (ChIP) assays, DNA electrophoretic mobility shift assays (EMSA), DNA pull-down assays, and microplate capture and detection assays. Commercial kits, materials, and reagents are available to practice many of these methods and, for example, can be obtained from the following suppliers: Thermo Scientific (Wilmington, DE), Signosis (Santa Clara, CA), Bio-Rad (Hercules, CA), and Promega (Madison, WI). A common approach to detect interactions between a polypeptide and a nucleic acid sequence is EMSA (see, e.g., Hellman L.M., et al., Nature Protocols 2:1849-1861 (2007)).

[00349] In another embodiment, the present invention includes a method of cutting a nucleic acid target sequences in a polynucleotide (e.g., a single-strand cut in dsDNA or double-strand cut in dsDNA) comprising providing one or more engineered Type I CRISPR-Cas effector complexes for introduction into a cell or biochemical reaction, and introducing the engineered Type I CRISPR-Cas effector complexes into the cell or biochemical reaction, thereby facilitating contact of the engineered Type I CRISPR-Cas effector complexes with the polynucleotide.

[00350] In one embodiment, a first engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a first nucleic acid target sequence in the polynucleotide and a first nuclease domain (e.g., FokI) (FIG. 15A, Cascade1, solid outline box, connected via a linker polypeptide, curved black line, to the first nuclease domain, represented as a circular sector), and a second engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a second nucleic acid target sequence in the polynucleotide and a second nuclease domain (e.g., FokI) (FIG. 15A, Cascade 2, dash outline box, connected via a linker polypeptide, curved black line, to the second nuclease domain, represented as a circular sector) are introduced into the cell or biochemical reaction. The first engineered Type I CRISPR-Cas

effector complex (FIG. 15B, Cascade1) binds to the first nucleic acid target sequence in dsDNA (FIG. 15B, dsDNA represented by paired, horizontal black lines) and the first nuclease domain cleaves the first strand of a dsDNA (FIG. 15C, Cascade1), and the second engineered Type I CRISPR-Cas effector complex (FIG. 15B, Cascade2) binds to the second nucleic acid target sequence in dsDNA and the second nuclease domain cleaves the second strand of a dsDNA. The binding of the engineered Type I CRISPR-Cas effector complexes results in cutting of the nucleic acid target sequences in the polynucleotide (e.g., a dsDNA) by the engineered Type I CRISPR-Cas effector complexes.

[00351] In an additional embodiment, a first engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a first nucleic acid target sequence in the polynucleotide, a second engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a second nucleic acid target sequence in the polynucleotide, and a Cas3 nickase (e.g., an ATPase-deficient Cas3 variant having only nickase activity) are introduced into the cell or biochemical reaction. The first engineered Type I CRISPR-Cas effector complex binds to the first nucleic acid target sequence in dsDNA, the Cas3 nickase protein associates with the first complex, and cleaves the first strand of a dsDNA, and the second engineered Type I CRISPR-Cas effector complex binds to the second nucleic acid target sequence in dsDNA, the Cas3 nickase protein associates with the second complex, and cleaves the second strand of a dsDNA. The binding of the engineered Type I CRISPR-Cas effector complexes with associated Cas3 nickase proteins results in cutting of the nucleic acid target sequences in the polynucleotide (e.g., a dsDNA) by the engineered Type I CRISPR-Cas effector complexes. Example 25A, Example 25B, and Example 25C present data that demonstrate Cascade RNP complexes comprising Cas3 ATPase deficient mutant proteins can induce targeted genomic deletions through paired nicking. This paired nicking can facilitate targeted deletions in the genomes of host cells (e.g., human cells).

[00352] In another embodiment, an engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a nucleic acid target sequence in the polynucleotide and a first nuclease domain (e.g., FokI) (FIG. 16A, Cascade; dash outline box, connected via a linker polypeptide, curved black line, to the first nuclease domain, represented as a circular sector), and a dCas3*-second nuclease domain (e.g., FokI) fusion protein (FIG. 16A, dCas3; solid outline box, connected via a linker polypeptide, curved black line, to the second nuclease domain, represented as a circular sector) capable of associating with the complex are introduced into the cell or biochemical reaction. The engineered Type I CRISPR-Cas effector complex (FIG. 16B, Cascade) binds to a nucleic acid target sequence in dsDNA (FIG. 16B, paired,

horizontal, black lines) and cleaves the first strand of a dsDNA (FIG. 16C, Cascade), and the dCas3* fusion protein associates with the Cascade RNP complex (FIG. 16B, dCas3*) and cleaves the second strand of the dsDNA (FIG. 16C, dCas3*).

[00353] In a further embodiment, an engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a target region comprising a nucleic acid target sequence in the polynucleotide and Cas3 protein (e.g., a Cas3 protein or a mCas3 protein) capable of associating with the complex are introduced into the cell or biochemical reaction. The engineered Type I CRISPR-Cas effector complex binds to a nucleic acid target sequence in dsDNA, the Cas3 protein (e.g., a Cas3 protein or a mCas3 protein) associates with the complex and cleaves at least one strand of the dsDNA in the target region. In some embodiments, cleavage of the dsDNA by the mCas3 protein results in a deletion in the target region of the dsDNA. This method can be used to make long range deletions of a specific length and can be useful for creation of gene knock-outs or knock-ins. In some embodiments, the Cas3 protein (e.g., a Cas3 protein or a mCas3 protein) can be fused to a Cascade complex subunit protein (e.g., a Cas7 protein, a Cas8 protein, a Cas5 protein, a Cse2 protein). Example 23A to Example 23C describe embodiments of mCas3 proteins.

[00354] In another embodiment, the present invention relates to using Type I CRISPR-Cas effector complexes, wherein a nuclease domain is fused to a Cascade complex protein (see, e.g., Example 11A, Table 38) or to a dCas3* protein (e.g., a dCas3* protein fused to a DNase) to delete nucleic acid target sequences. This method can be used to make cuts as well as deletions in a target region of dsDNA and can be useful for creation of gene knock-outs. In some embodiments, the nuclease domain can be fused to a Cascade complex subunit protein such as a Cas7 protein, a Cas8 protein, a Cas5 protein, a Cse2 protein.

[00355] Methods of cutting a nucleic acid target sequence in a polynucleotide can further comprise introduction of a donor polynucleotide into a cell to facilitate incorporation of at least a portion of the donor polynucleotide into gDNA of the cell.

[00356] FIG. 17A illustrates an example of both strands of a dsDNA (FIG. 17A, paired, dark horizontal lines) being cleaved by a first engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a first nucleic acid target sequence in the polynucleotide (FIG. 17A, Cascade1) and a first nuclease domain (e.g., FokI) (FIG. 17A, linker polypeptide illustrated as a bent line connecting Cascade1 and a grey, circular sector), and a second engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a second nucleic acid target sequence in the polynucleotide (FIG. 17A, Cascade 2) and a second nuclease domain (e.g., FokI) (FIG. 17A, linker polypeptide illustrated as a bent line connecting

Cascade2 and a grey, circular sector). FIG. 17B illustrates a donor polynucleotide (FIG. 17B, paired, dashed lines shown above Cascade2) comprising homology arms complementary to DNA sequences adjacent the double-strand cut site (FIG.18B, Donor, dashed lines). FIG. 17C illustrates incorporation of a portion of the donor polynucleotide (FIG. 17C, paired, dashed lines connecting the paired, dark, horizontal lines representing dsDNA) in the region of the double-strand cut site. Incorporation of the donor polynucleotide is mediated by cellular DNA repair mechanisms (e.g., HDR) (FIG. 17B to FIG. 17C, downward pointing, vertical arrow represents cellular DNA repair mechanisms).

[00357] In other embodiments, an engineered Type I CRISPR-Cas effector complex comprising a guide complementary to a first nucleic acid target sequence in a polynucleotide and a first nuclease domain can be paired with a second component comprising a second nuclease domain, wherein the second component is capable of binding to a second nucleic acid target sequence in the polynucleotide. Examples of such second components include, a transcription activator-like effector nuclease (TALEN) comprising the second nuclease domain, a zinc finger nuclease (ZFN)comprising the second nuclease domain, or a dCas9/NATNA complex comprising the second nuclease domain.

[00358] In one embodiment, a region of a target polynucleotide (e.g., gDNA) can be deleted using a combination of a Cascade complex comprising a guide complementary to a first nucleic acid target sequence in the target polynucleotide and a dCas9/NATNA complex wherein the NATNA comprises a spacer sequence complementary to a second nucleic acid target sequence in the target polynucleotide. The first and second nucleic acid target sequences are selected to flank the nucleic acid target sequence targeted for deletion. A Cas3 protein comprising an active endonuclease activity associates with the Cascade complex and then progressively deletes a single strand of the dsDNA comprising the nucleic acid target sequence targeted for deletion. When the Cas3 protein collides with the dCas9/NATNA complex (i.e., a "roadblock"), the Cas3 nuclease activity can be stopped at the second nucleic acid target sequence by the dCas9/NATNA complex. FIG. 21A to FIG. 21D illustrate an example of a Cas3 deletion of a nucleic acid target sequence. FIG. 21A shows a dsDNA (FIG. 21A, paired, horizontal, black lines) comprising nucleic acid target sequence 1 (FIG. 21A, NATS1) and nucleic acid target sequence 2 (FIG. 21A, NATS2) that flank the nucleic acid target sequence targeted for deletion. FIG. 21A shows the Cascade complex comprising a guide complementary to NATS1 (FIG. 21A, Cascade; black line framed rectangle), the Cas3 protein (FIG. 21A, Cas3; grey circular sector), and the dCas9/NATNA complex comprising a spacer complementary to NATS2 (FIG. 21A, dCas9; dashed line framed rectangle). FIG. 21B shows binding of the

Cascade complex to NATS1, association of the Cas3 protein with the Cascade complex, and binding of the dCas9/NATNA complex to NATS2. FIG. 21C illustrates the progressive deletion by Cas3 of a single strand of the nucleic acid target sequence targeted for deletion. FIG. 21D shows the dissociation of the Cas3 protein from the dsDNA at the position of the dCas9/NATNA complex bound to NATS2. Example 24A to Example 24D present data that support the use of protein roadblocks to control the length of deletions mediated by Cas3 protein associated with Cascade nucleoprotein complexes; thus, providing a method to use Cas3 protein associated with Cascade nucleoprotein complexes to facilitate formation of deletions having a defined length in the gDNA of cells (e.g., human cells).

[00359] In another embodiment, a region of a target polynucleotide (e.g., gDNA) can be deleted using a combination of a first Cascade complex comprising a guide complementary to a first nucleic acid target sequence in the target polynucleotide and a second Cascade complex comprising a guide complementary to a second nucleic acid target sequence in the target polynucleotide. The first and second nucleic acid target sequences are selected to flank the nucleic acid target sequence targeted for deletion. Cas3 proteins comprising active endonuclease activity associate with each Cascade complex and then progressively delete both strands of the nucleic acid target sequence targeted for deletion. When each Cas3 protein collides with one of the Cascade complexes, the Cas3 nuclease activity can be stopped at the first and second nucleic acid target sequences by the Cascade complexes. FIG. 22A to FIG. 22D illustrate an example of a Cas3 deletion of both strands of a nucleic acid target sequence. FIG. 22A shows a dsDNA (FIG. 22A; paired, horizontal, black lines) comprising nucleic acid target sequence 1 (FIG. 22A, NATS1) and nucleic acid target sequence 2 (FIG. 22A, NATS2) that flank the nucleic acid target sequence targeted for deletion. FIG. 22A shows the first Cascade complex comprising a guide complementary to NATS1 (FIG. 22A, Cascade1; black line framed rectangle), the Cas3 proteins (FIG. 22A, Cas3; grey circular sector), and the second Cascade complex comprising a guide complementary to NATS2 (FIG. 22A, Cascade2; dash line framed rectangle). FIG. 22B shows binding of the Cascade complexes to NATS1 and NATS2, as well as association of the Cas3 proteins with the Cascade complexes. FIG. 22C illustrates the progressive deletion resulting from movement along the DNA and nuclease degradation by Cas3 of both strands of the nucleic acid target sequence targeted for deletion. FIG. 22D shows the dissociation of the Cas3 proteins from the dsDNA at the positions of the Cascade complexes bound to NATS1 and NATS2.

[00360] In a further embodiment, a Cascade complex can be modified such that it is not capable of binding to a Cas3 protein, and such a modified Cascade complex can act as a

roadblock essentially in the same manner as illustrated in FIG. 21A to FIG. 21D to stop progressive degradation of DNA by a catalytically active Cas3 in association with a Cascade RNP complex. Additional site-specific binding proteins (e.g., transcription activator-like effectors (TAL), or zinc-finger (ZnF) DNA binding proteins) can be used as roadblocks in a similar way.

[00361] In some embodiments, the nucleic acid target sequence is dsDNA (e.g., genomic) DNA. In some embodiments, the nucleic acid target sequence is double-stranded and one or both of the strands is cut. Such methods of cutting a nucleic acid target sequence can be carried out *in vitro*, *in vivo*, or *ex vivo*.

[00362] As described above, in some embodiments the present invention relates to introducing one or more engineered Type I CRISPR-Cas effector complexes into a host cell to facilitate cleavage of a nucleic acid target sequence in dsDNA in the presence of a donor polynucleotide, wherein the one or more engineered Type I CRISPR-Cas effector complexes generate a cut site (or cut site and associate deletion) in a target region comprising the nucleic acid target sequence of the host cell DNA thereby facilitating insertion of at least a portion of the donor polynucleotide into the target region. In some embodiments, the cut site is a double-stranded break in the target region (e.g., when using two engineered Type I CRISPR-Cas effector complexes each comprising a spacer and a fusion protein comprising a Cas protein and an endonuclease (e.g., a FokI) or two engineered Type I CRISPR-Cas effector complexes each comprising a spacer that associate with a Cas3 protein or a mCas3 protein). In some embodiments, the cut site is a single-stranded break in the target region (e.g., when using a Type I CRISPR-Cas effector complex associated with a mCas3 protein). In other embodiments, the cut site is a deletion in the target region (e.g., when using a Type I CRISPR-Cas effector complex associated with a Cas3 or mCas3 protein).

[00363] In order to demonstrate homology directed repair (HDR), a minimal CRISPR array was designed to target the FokI-PseCascade RNP complex to four loci (WDR92, B2M, CCR5, and TRAC) in the human genome. The minimal CRISPR arrays were generated with PCR-based assembly using three oligonucleotides (SEQ ID NO:1513 to SEQ ID NO:1515; Example 20A) and a unique primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence, wherein the first and second spacers directed FokI-PseCascade RNP complexes to adjacent nucleic acid target sequences to enable FokI dimerization and genome cleavage (i.e., generation of a cut site).

[00364] For each HDR insertion site in a target region comprising a cut site, in this case overlapping with the cut site, cells were transfected with the following: 3 μg of vector encoding

FokI-PseCascade complex protein components comprising FokI fused to the N-terminus of Cas8 with an NLS connected to the N terminus of the FokI, 150 ng of the minimal CRISPR arrays, and 0-60 pmol of a single-stranded oligodeoxynucleotide (ssODN) template donor polynucleotide for HDR. The ssODN comprised homology arms, each homology arm was 75 nucleotides, and the two arms were symmetrically located around the cut site. The donor polynucleotide further comprised phosphorothioate bonds at the 3' terminal nucleotides of homology arms to reduce or prevent cellular degradation of the donor polynucleotide. 5' of the phosphothiorate bonds, the donor polynucleotide further comprised an insertion sequence of "TAATAAT" to insert two stop codons, and increase the interspacer distance in the repaired chromosome, thus impeding FokI-PseCascade RNP complex re-cleavage.

[00365] Transfection was carried out in HEK293 cells essentially as described in Example 20B with the exception that a ssODN was included in the mixture to enable HDR. Several days after transfection, gDNA was purified from the cells, treated with exonuclease to remove any residual ssODN that could contaminate subsequent PCRs, and then used as template for amplification to measure donor insertion. Deep sequencing analysis was carried out essentially as described in Example 20C. The percentage of mutant reads out of total reads from this experiment are presented in Table 13 (the first column is pmol of ssODN):

[00366]

Table 13
Percentage of Mutant Reads

	WDR92			**B2M**			**CCR5**			**TRAC**		
	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3
0	11.6	9.54	9.93	3.46	3.22	5.88	7.02	4.94	8.75	3.76	2.94	3.38
20	11	8.82	12.6	0.99	4.99	2.94	5.56	6.69	12.9	2.92	3.72	3.8
40	10.1	9.54	11.3	1.97	2.79	0	5.44	0.24	5.13	3.69	3.99	4.97
60	12.6	10.4	11.5	0.94	1.01	1.08	n/a	5.95	8.5	4	3.48	n/a

[00367] The percentage of mutant reads indicates the mutant reads that contain indels resulting from non-homologous end-joining as well as insertions of the "TAATAAT" HDR sequence.

[00368] The percentage of HDR reads, containing only the "TAATAAT" insertion sequence, out of total mutants reads from this experiment are presented in Table 14 (the first column is pmol of ssODN):

[00369]

Table 14
Percentage of HDR Reads

	WDR92			B2M			CCR5			TRAC		
	Rep1	Rep2	Rep3	Rep1	Rep2	Rep3	Rep 1	Rep2	Rep3	Rep1	Rep 2	Rep3
0	0	0	0	0	0	0	0	0	0	0	0	n/a
20	23.03	10.72	9.68	14.26	0.11	14.38	7.07	3.94	0.22	21.37	0	9.14
40	16.38	13.7	4.83	0	16.5	0	2.42	0	10.54	2.1	0	13.23
60	5.5	11.98	7.31	0	12.28	5	n/a	1.9	13.87	0.79	7.53	9.38

[00370] As can be seen from the data, cleavage of dsDNA by Cascade RNP complexes enables HDR and incorporation of donor polynucleotide at multiple loci across the human genome.

[00371] In yet another embodiment, the present invention includes a method of modifying one or more nucleic acid target sequences in a polynucleotide (e.g., DNA) in a cell or biochemical reaction comprising providing one or more engineered Type I CRISPR-Cas effector complexes (e.g., comprising a Cas subunit protein-cytidine deaminase fusion protein) for introduction into the cell or the biochemical reaction, and introducing the engineered Type I CRISPR-Cas effector complex(es) into the cell or biochemical reaction, thereby facilitating contact of the engineered Type I CRISPR-Cas effector complex(es) with the polynucleotide resulting in binding of the engineered Type I CRISPR-Cas effector complex(es) to the nucleic acid target sequence(s) in the polynucleotide that facilitates mutation of the nucleic acid target sequence(s) (e.g., C-to-T, G-to-A, A-to-G, and T-to-C). FIG. 18A to FIG. 18D illustrate an example of using a Cascade complex comprising a Cas subunit protein-linker polypeptide-cytidine deaminase fusion protein (Cascade/CD complex) to mutate a target nucleotide in gDNA of a cell (FIG. 18A, paired, dark horizontal lines, with a "C" for cytosine and a "G" for guanine). The Cascade/CD complex (FIG. 18A; "Cascade" with linker polypeptide illustrated as a bent line connecting Cascade and the cytidine deaminase "CD" represented as a grey, circular sector) is introduced into the cell. The Cascade/CD complex comprises a guide complementary to a DNA target sequence adjacent a target cytosine (FIG. 18B, "C"). In FIG. 18B, the Cascade/CD complex binds the DNA target sequence and the cytidine deaminase converts the cytosine (FIG. 18B, "C") to a uracil (FIG. 18C, "U"). Cellular repair mechanisms can then repair the uracil to a thymidine, and change the mismatched guanidine to adenine

(FIG. 18C to FIG. 18D, downward pointing, vertical arrow represents cellular DNA repair mechanisms).

[00372] In yet another embodiment, the present invention includes methods of modulating *in vitro* or *in vivo* transcription, for example, transcription of a gene comprising regulatory element sequences. Such methods comprise providing one or more engineered Type I CRISPR-Cas effector complexes (e.g., comprising a Cas subunit protein-transcription factor fusion protein) for introduction into the cell or the biochemical reaction, and introducing the engineered Type I CRISPR-Cas effector complex(es) into the cell or biochemical reaction, thereby facilitating contact of the engineered Type I CRISPR-Cas effector complex(es) with the regulatory element sequences resulting in binding of the engineered Type I CRISPR-Cas effector complex(es) to the regulatory element sequences thereby facilitating modulating *in vitro* or *in vivo* transcription of the gene comprising the regulatory element sequences.

[00373] FIG. 19A and FIG. 19B present general illustrations of examples for the transcriptional activation of a generic gene ("GENE1"). FIG. 19A provides an overview of transcriptional regulation of an endogenous gene in a eukaryotic cell. In FIG. 19A, the two dark parallel lines represent double-stranded DNA, the location of Gene 1 (FIG. 19A, GENE 1) is indicated, as well as the transcriptional start site (FIG. 19A, TSS) associated with Gene 1. In the first panel of FIG. 19A, a transcription factor (FIG. 19A, TF) that is needed for the transcriptional activation of Gene 1 and polymerase II (FIG. 19A, Pol II) are illustrated as not yet associated with Gene1-TSS. The second panel illustrates association of the TF with its cognate TSS. The TF then recruits a transcription activation protein (FIG. 19A, TP) that then recruits RNA polymerase II (FIG. 19A, Pol II). Typically, in eukaryotes the TF factor and the TP form a complex comprising multiple proteins and possibly other molecules. The third panel illustrates the resulting transcription of Gene 1 by Pol II (FIG. 19A, bent arrow at the end of GENE 1 indicates the direction of transcription). This type of transcriptional activation is typically dependent on TF(s) that are specific to the expression of a gene(s). FIG. 19B presents an illustration of one embodiment of the present invention, wherein a Cascade complex is engineered to comprise a protein or factor (FIG. 19B, CASCADEa) that attracts one or more components in the cells responsible for transcriptional activation (Transcriptional Activation factor; FIG. 19B, TA). An example of one such protein or factor is the protein VP64. CASCADEa comprises a guide that is capable of binding at or near the TSS (FIG. 19B, TSS). In FIG. 19B, the two dark parallel lines represent double-stranded DNA, the location of Gene 1 (FIG. 19B, GENE 1) is indicated, as well as the transcriptional start site (TSS) associated with Gene 1. In the first panel of FIG. 19B, CASCADEa and polymerase II (FIG. 19B, Pol II) are

illustrated as not yet associated with Gene1-TSS. The second panel illustrates association of CASCADEa with its target, the TSS. The CASCADEa then recruits a transcription activation protein (FIG. 19B, TA) that then recruits RNA polymerase II (FIG. 19B, Pol II). The third panel illustrates the resulting transcription of Gene 1 by Pol II (FIG. 19B, bent arrow at the end of GENE 1 indicates the direction of transcription). One advantage of this embodiment of the present invention is that transcriptional activation of a gene is not dependent on endogenous transcription factors that bind to the TSS of the gene, rather the TSS of a gene can be targeted by selection of an appropriate Cascade guide.

[00374] FIG. 20A and FIG. 20B present a general illustration of an example for the transcriptional repression of a generic gene (FIG. 20A, GENE 1) using a Cascade complex comprising a Cas subunit protein-KRAB domain fusion and a guide (FIG. 20A, CASCADEi with linker polypeptide illustrated as a bent line connecting Cascade and a circular element representing a KRAB domain) complementary to regulatory sequences (FIG. 20A, promoter) associated with GENE 1. Binding of CASCADEi to the regulatory sequences (FIG. 20B) results in transcriptional repression of GENE 1 (FIG. 20B, dark line ending in X represents transcriptional repression).

[00375] The engineered Type I CRISPR-Cas effector complexes, as described herein, can be incorporated into a kit. In some embodiments, a kit includes a package with one or more containers holding the kit elements, as one or more separate compositions or, optionally if the compatibility of the components allows, as admixture. In some embodiments, a kit also comprises one or more of the following excipients: a buffer, a buffering agent, a salt, a sterile aqueous solution, a preservative, and combinations thereof. Illustrative kits can comprise one or more engineered Type I CRISPR-Cas effector complexes and one or more excipients, or one or more nucleic acid sequences encoding one or more components of engineered Type I CRISPR-Cas effector complexes.

[00376] Furthermore, kits can further comprise instructions for using engineered Type I CRISPR-Cas effector complex compositions.

[00377] Another aspect of the invention relates to methods of making or manufacturing one or more engineered Type I CRISPR-Cas effector complexes, or components thereof. In one embodiment, a method of making or manufacturing comprises production of engineered Type I CRISPR-Cas effector complexes in a cell and purification of the engineered Type I CRISPR-Cas effector complexes from cell lysates.

[00378] Engineered Type I CRISPR-Cas effector complex compositions can further comprise a detectable label, such as a moiety that can provide a detectable signal. Examples of

detectable labels include, but are not limited to, an enzyme, a radioisotope, a member of a specific binding pair, a fluorophore (FAM), a fluorescent protein (green fluorescent protein (GFP), red fluorescent protein, mCherry, tdTomato), a DNA or RNA aptamer together with a suitable fluorophore (enhanced GFP (eGFP), "Spinach"), a quantum dot, an antibody, and the like. A large number and variety of suitable detectable labels are well-known to one of ordinary skill in the art.

[00379] In some embodiments, engineered Type I CRISPR-Cas effector complexes (i.e., nucleoprotein particles) can be introduced into cells by methods including, but not limited to, nucleofection, gene gun delivery, sonoporation, cell squeezing, lipofection, or the use of other chemicals, cell penetrating peptides, and the like. In other embodiments, coding sequences for one or more components of engineered Type I CRISPR-Cas effector complexes and associated proteins can be introduced into cells using vector systems, expression cassettes comprising DNA sequences encoding one or more of the components, as well as one or more RNA molecules (e.g., mRNA) comprising expression cassettes comprising RNA sequences encoding one or more of the components.

[00380] One embodiment of the present invention relates to the use of engineered Type I CRISPR-Cas effector complexes to produce recombinant cells (e.g., modified lymphocytes). The method typically comprises facilitating contact of a dsDNA, comprising a target region comprising a nucleic acid target sequence in a host cell with one or more engineered Class 1 Type I CRISPR-Cas effector complexes of the present invention. Contact of the engineered Class 1 Type I CRISPR-Cas effector complex with the nucleic acid target sequence results in binding of the engineered Class 1 Type I CRISPR-Cas effector complex with the target region comprising the nucleic acid target sequence, cleavage of the target region comprising the nucleic acid target sequence, and modification of the dsDNA in the target region, thus producing the recombinant cell. In some embodiments, the dsDNA comprises more than one nucleic acid target sequence and engineered Class 1 Type I CRISPR-Cas effector complexes comprising spacer sequences complementary to each nucleic acid target sequence are used to bind, cut, and modify each nucleic acid target sequence. In some embodiments, the modification of the target region is an insertion, a deletion, or insertion and deletion. Methods of cutting a nucleic acid target sequences in a polynucleotide (e.g., a single-strand cut in dsDNA or double-strand cut in dsDNA) comprising providing one or more engineered Type I CRISPR-Cas effector complexes for introduction into a cell are described above.

[00381] Embodiments of the present invention include producing recombinant cells using one or more engineered Class 1 Type I CRISPR-Cas effector complexes, wherein the

gDNA of the recombinant cells comprise knock-out mutations (e.g., of the B2M gene and/or PDCD1 gene), knock-ins (e.g., editing at the TRAC locus and integration of a CAR from a donor polynucleotide), or combinations thereof. In some embodiments, cleavage at a nucleic acid target sequence in a TRAC gene of gDNA is followed by incorporation of at least a portion of a donor polynucleotide at the nucleic acid target sequence. The donor polynucleotide can comprise a CAR construct, wherein the CAR is inserted in the nucleic acid target sequence.

[00382] Recombinant cells made by methods of the present invention can be used in adoptive cell transfer (ACT). ACT is a rapidly emerging immunotherapy approach that uses transplanted immune cells to treat cancer. ACT is the transfer of cells into a patient. Most commonly, the immune cells are derived from the immune system with the goal of improving immune function. In autologous cancer immunotherapy, immune cells or stem cells are harvested from a patient and expanded by culturing *ex vivo* to large quantities and then returned to the patient. The immune cells or stem cells can be modified in a variety of ways in culture (e.g., use of genome editing to incorporate a CAR into the genome of a T cell). In some embodiments, lymphocytes for modification are isolated from a subject, modified, and then reintroduced into the same subject. This technique is known as autologous lymphocyte therapy. In allogeneic cancer immunotherapy, culture expanded immune cells or stem cells originating from a single donor provide treatments to large numbers of patients. Such immune cells or stems cells can also be modified in a variety of ways in culture. In some embodiments, lymphocytes can be isolated, modified, and introduced into a different subject. This technique is known as allogenic lymphocyte therapy.

[00383] In certain embodiments, such immunotherapy methods can utilize lymphocytes including but not limited a T cell, a natural killer cell (NK cell), a B cell, a tumor infiltrating lymphocyte (TIL), a chimeric antigen receptor T cell (CAR-T cell), a T cell receptor engineered T cell (TCR), a TCR CAR-T cell, a CAR TIL cell, a CAR-NK cell, an engineered NK cell, or a hematopoietic stem cell that gives rise to a lymphocyte cell. In other embodiments, the cell is a stem cell, a dendritic cell, or the like. The genomes of such cells can be modified (e.g., generation of insertions and/or deletions in a lymphocyte cell genome) by use of one or more engineered Class I Type I Cascade effector complexes of the present invention.

[00384] Lymphocytes for modification can be isolated from a subject, such as a human subject, for example from blood or from solid tumors, such as in the case of TILs, or from lymphoid organs such as the thymus, bone marrow, lymph nodes, and mucosal-associated lymphoid tissues. Techniques for isolating lymphocytes are well known in the art. For example, lymphocytes can be isolated from peripheral blood mononuclear cells (PBMCs), which are

separated from whole blood using, for example, ficoll, a hydrophilic polysaccharide that separates layers of blood, and density gradient centrifugation. Generally, anticoagulant or defibrinated blood specimens are layered on top of a ficoll solution, and centrifuged to form different layers of cells. The bottom layer includes red blood cells (erythrocytes), which are collected or aggregated by the ficoll medium and sink completely through to the bottom. The next layer contains primarily granulocytes, which also migrate down through the ficoll-paque solution. The next layer includes lymphocytes, which are typically at the interface between the plasma and the ficoll solution, along with monocytes and platelets. To isolate the lymphocytes, this layer is recovered, washed with a salt solution to remove platelets, ficoll and plasma, then centrifuged again. Alternatively, cells can be isolated from donor blood through centrifugation techniques (e.g., using a CellSaver® (Haemonetrics, Braintree, MA) machine or a Lovo Automated Cell Processing System (Fresenius Kabi USA, LLC, Lake Zurich, IL)).

[00385] Other techniques for isolating lymphocytes include biopanning, which isolates cell populations from solution by binding cells of interest to antibody-coated plastic surfaces. Unwanted cells are then removed by treatment with specific antibody and complement. Additionally, fluorescence-activated cell sorting (FACS) analysis can be used to detect and count lymphocytes. FACS analysis uses a flow cytometer that separates labelled cells based on differences in light scattering and fluorescence.

[00386] For TILs, lymphocytes are isolated from a tumor and grown, for example, in high-dose IL-2 and selected using cytokine release coculture assays against either autologous tumor or HLA-matched tumor cell lines. Cultures with evidence of increased specific reactivity compared to allogeneic non-MHC matched controls are selected for rapid expansion and then introduced into a subject in order to treat cancer (see, e.g., Rosenberg, S., et al., Clin. Cancer Res. 17:4550-4557 (2011); Dudly, M., et al., Science 298:850-854 (2002); Dudly, M., et al., J. Clin. Oncol. 26:5233-5239 (2008); Dudley, M., et al., J. Immnother. 26:332-342 (2003)).

[00387] Upon isolation, lymphocytes can be characterized in terms of specificity, frequency and function. Frequently used assays include an ELISPOT assay, which measures the frequency of T cell response.

[00388] In some embodiments, CD4+ and CD8+ T cells are isolated from donor peripheral blood mononuclear cell (PBMCs). One of ordinary skill in the art can isolate T-cells, or other lymphoid cells, by a variety of methods as described above. Such cells also can be isolated by differentiation from iPSC cells.

[00389] After isolation, lymphocytes can be activated using techniques known in the art in order to promote proliferation and differentiation into specialized effector lymphocytes.

Surface markers for activated T cells include, for example, CD3, CD4, CD8, PD1, IL2R, and others. Activated cytotoxic lymphocytes can kill target cells after binding cognate receptors on the surface of target cells. Surface markers for NK cells include, for example CD16, CD56, and others.

[00390] Following isolation and optionally activation, lymphocytes can be modified in order to provide desired characteristics. One or more engineered Type I Cascade effector complexes of the present invention can be used to introduce genomic modifications including, but not limited to, introduction of coding sequences to be expressed and/or inactivating endogenous gene expression. In some embodiments, one or more engineered Type I Cascade effector complexes of the present invention can be used for editing the TRAC gene (encoding T cell receptor α constant), B2M gene (encoding β2 microglobulin), and/or PDCD1 gene (encoding programmed cell death protein 1; also known as PD-1).

[00391] T cells and NK cells are examples of lymphocytes that can be modified by the methods of the present invention. In some embodiments, one or more engineered Type I Cascade effector complexes of the present invention can be used to introduce a cut site in a target region of a gene in the presence of a donor polynucleotide comprising a CAR, wherein at the CAR is incorporated into the target region of the genome of the lymphocyte. In additional embodiments, one or more engineered Type I Cascade effector complexes of the present invention can be used to introduce a cut site in a target region of a gene to facilitate generation of a knock-out mutation to prevent expression of the gene.

[00392] In another embodiment engineered Type I Cascade effector complexes of the present invention can be used to introduce genomic modifications into human iPSCs. In some embodiments, one or more engineered Type I Cascade effector complexes of the present invention can be used for editing the TRAC gene, the B2M gene, and/or the PDCD1gene. In further embodiments, engineered Type I Cascade effector complexes together with a donor polynucleotide can be used to introduce genomic modifications and coding sequences such as a CAR or a cytokine (e.g., IL2, IL15 and the like). The modified iPSC cells can then be differentiated further to mature cell types comprising T cells and NK cells or dendritic cells. In some embodiments, modified iPSCs can be differentiated to CAR-T cells and CAR-NK cells.

[00393] In some embodiments of the methods of the present invention, the donor polynucleotide comprises a polynucleotide encoding a CAR. The CAR can be targeted for insertion into a target region of a gene (e.g., the TRAC gene) comprising a cut site via homologous recombination ("knock-in"). An advantage of this approach is that it can also provide a knock-out of the targeted TRAC gene; that is, the TRAC gene is rendered

inoperative. Example of extracellular antigen-recognizing domains that can be incorporated into CAR constructs are described above (see Table 2). In one embodiment, the extracellular antigen-recognizing domain comprises a CD19 binding moiety (e.g., an anti-CD19 scFv). In another embodiment, the extracellular antigen-recognizing domain comprises a B-cell maturation antigen (BCMA) binding moiety (e.g., an anti-BCMA scFv).

[00394] In embodiments of the methods of the present invention comprising generating a cut site in a target region of DNA, the method can further comprise introducing a donor polynucleotide into the modified cell thereby facilitating insertion of at least a portion of the donor polynucleotide into the target region comprising a cut site of the modified cell. The donor polynucleotide can be directly introduced into the modified cell. In some embodiments, the donor polynucleotide is introduced in using a vector. General methods for the construction of vectors are known in the art. Examples of viral vectors include, but are not limited to, lentivirus, retrovirus, adenovirus, herpes simplex virus I or II, parvovirus, reticuloendotheliosis virus, and AAV vectors.

[00395] A further embodiment of the methods of the present invention comprises introduction of a mutation in the B2M gene. In preferred embodiments, the mutation is a knock-out mutation in the B2M gene.

[00396] A further embodiment of the methods of the present invention comprises introduction of a mutation in the PDCD1 gene. In preferred embodiments, the mutation is a knock-out mutation in the PDCD1 gene.

[00397] Genomic modifications facilitated by one or more engineered Type I Cascade effector complexes of the present invention can be performed by simultaneously or serially introducing either engineered Cascade complexes, polynucleotides (e.g, plasmids or expression cassettes), or mixtures thereof into a host cell (e.g., a lymphocyte).

[00398] After producing the modified lymphocytes, the lymphocytes can be screened to select for cells either expressing (e.g., expressing the desired cell surface receptor) or not expressing (e.g., a cell surface protein whose expression has been inactivated through genome editing using one or more engineered Type I Cascade effector complexes), using methods such as high-throughput screening techniques including, but not limited to, FACS, microfluidics-based screening platforms, and the like. These techniques are known in the art (see, e.g., Wojcik, M., et al., Int. J. Mol.. Sci. 16:24918-24945 (2015)).

[00399] Once produced, the modified lymphocytes can be formulated into pharmaceutical compositions for delivery to the subject to be treated. Compositions of the present invention include a modified lymphocyte and one or more pharmaceutically acceptable

excipients. Exemplary excipients include, without limitation, carbohydrates, inorganic salts, antimicrobial agents, antioxidants, surfactants, buffers, acids, bases, and combinations thereof. Excipients suitable for injectable compositions include water, alcohols, polyols, glycerine, vegetable oils, phospholipids, and surfactants. A carbohydrate such as a sugar, a derivatized sugar such as an alditol, aldonic acid, an esterified sugar, and/or a sugar polymer may be present as an excipient. Specific carbohydrate excipients include, for example: monosaccharides, such as fructose, maltose, galactose, glucose, D-mannose, sorbose, and the like; disaccharides, such as lactose, sucrose, trehalose, cellobiose, and the like; polysaccharides, such as raffinose, melezitose, maltodextrins, dextrans, starches, and the like; and alditols, such as mannitol, xylitol, maltitol, lactitol, xylitol, sorbitol (glucitol), pyranosyl sorbitol, myoinositol, and the like. The excipient can also include an inorganic salt or buffer such as citric acid, sodium chloride, potassium chloride, sodium sulfate, potassium nitrate, sodium phosphate monobasic, sodium phosphate dibasic, and combinations thereof. Freezing agents (e.g., CryoStor® (BioLife Solutions Inc, Bothell, WA) CS2, CS5, or CS10 freeze media) can be used to freeze cells for storage and transport.

[00400] A pharmaceutical composition of the invention can also include an antimicrobial agent for preventing or deterring microbial growth. Nonlimiting examples of antimicrobial agents suitable for the present invention include benzalkonium chloride, benzethonium chloride, benzyl alcohol, cetylpyridinium chloride, chlorobutanol, phenol, phenylethyl alcohol, phenylmercuric nitrate, thimerosal, and combinations thereof.

[00401] An antioxidant can also be present in the pharmaceutical composition. Antioxidants are used to prevent oxidation, thereby preventing the deterioration of the lymphocytes or other components of the preparation. Suitable antioxidants for use in the present invention include, for example, ascorbyl palmitate, butylated hydroxyanisole, butylated hydroxytoluene, hypophosphorous acid, monothioglycerol, propyl gallate, sodium bisulfite, sodium formaldehyde sulfoxylate, sodium metabisulfite, and combinations thereof.

[00402] A surfactant can be present as an excipient. Exemplary surfactants include: polysorbates, such as TWEEN 20 and TWEEN 80, and pluronics such as F68 and F88 (BASF, Mount Olive, New Jersey); sorbitan esters; lipids, such as phospholipids such as lecithin and other phosphatidylcholines, phosphatidylethanolamines (although preferably not in liposomal form), fatty acids and fatty esters; steroids, such as cholesterol; chelating agents, such as EDTA; and zinc and other such suitable cations.

[00403] Acids or bases can be present as an excipient in the pharmaceutical composition. Nonlimiting examples of acids that can be used include those acids selected from the group

consisting of hydrochloric acid, acetic acid, phosphoric acid, citric acid, malic acid, lactic acid, formic acid, trichloroacetic acid, nitric acid, perchloric acid, phosphoric acid, sulfuric acid, fumaric acid, and combinations thereof. Examples of suitable bases include, without limitation, bases selected from the group consisting of sodium hydroxide, sodium acetate, ammonium hydroxide, potassium hydroxide, ammonium acetate, potassium acetate, sodium phosphate, potassium phosphate, sodium citrate, sodium formate, sodium sulfate, potassium sulfate, potassium fumerate, and combinations thereof.

[00404] The amount of lymphocytes (or other recombinant cells) in the composition will vary depending on a number of factors, but will optimally be a therapeutically effective dose when the composition is in a unit dosage form or container (e.g., a bag). A therapeutically effective dose can be determined experimentally by repeated administration of increasing amounts of the composition in order to determine which amount produces a clinically desired endpoint.

[00405] The amount of any individual excipient in the composition will vary depending on the nature and function of the excipient and particular needs of the composition. Typically, the optimal amount of any individual excipient is determined through routine experimentation, i.e., by preparing compositions containing varying amounts of the excipient (ranging from low to high), examining the stability and other parameters, and then determining the range at which optimal performance is attained with no significant adverse effects. Generally, however, the excipient(s) will be present in the composition in an amount of about 1% to about 99% by weight, preferably from about 5% to about 98% by weight, more preferably from about 15% to about 95% by weight of the excipient, with concentrations less than 30% by weight most preferred. These foregoing pharmaceutical excipients along with other excipients are described in "Remington: The Science & Practice of Pharmacy," current edition, Williams & Williams; the "Physician's Desk Reference," current edition, Medical Economics, Montvale, NJ; and Kibbe, A.H., Handbook of Pharmaceutical Excipients, current edition, American Pharmaceutical Association, Washington, D.C.

[00406] The pharmaceutical compositions can be housed in a syringe, an implantation device, or the like, depending upon the intended mode of delivery and use. Preferably, the amount of the composition present is appropriate for a single dose, in a premeasured or pre-packaged form.

[00407] The pharmaceutical compositions herein may optionally include one or more additional agents, such as other medications used to treat a subject for the cancer in question or to treat known side-effects from the treatment. For example, T cells release cytokines into the

bloodstream, which can lead to dangerously high fevers and precipitous drops in blood pressure. This condition is known as cytokine release syndrome (CRS). In many patients, CRS can be managed with standard supportive therapies, including steroids and immunotherapies, such as tocilizumab (Actemra™, Genentech, South San Francisco, CA) that block IL-6 activity.

[00408] At least one therapeutically effective cycle of treatment with a modified lymphocyte composition will be administered to a subject. By "therapeutically effective cycle of treatment" is intended a cycle of treatment that, when administered, brings a positive therapeutic response with respect to treatment of an individual for the disease in question. By "positive therapeutic response" is intended that the individual undergoing treatment according to the invention exhibits an improvement in one or more symptoms of the disease, including such improvements as tumor reduction and/or reduced need for lymphocyte therapy.

[00409] In certain embodiments, multiple therapeutically effective doses of compositions comprising the lymphocytes or other medications will be administered. The compositions of the present invention are typically, although not necessarily, administered via injection, such as subcutaneously, intradermally, intravenously, intraarterially, intramuscularly, intraperitoneally, intramedullary, intratumorally, intranodally), by infusion, or locally. The pharmaceutical preparation can be in the form of a liquid solution or suspension immediately prior to administration. The foregoing is meant to be exemplary as additional modes of administration are also contemplated. The pharmaceutical compositions may be administered using the same or different routes of administration in accordance with any medically acceptable method known in the art.

[00410] The actual dose to be administered will vary depending upon the age, weight, and general condition of the subject as well as the severity of the condition being treated, the judgment of the health care professional, and particular lymphocytes being administered. Therapeutically effective amounts can be determined by those skilled in the art, and will be adjusted to the particular requirements of each particular case.

[00411] Generally, a therapeutically effective amount of lymphocytes will range from a total of about 1×10^5 to about 1×10^{10} lymphocytes or more per patient, such as 1×10^6 to about 1×10^{10}, e.g., 1×10^7 to 1×10^9, such as 5×10^7 to 5×10^8, or any amount within these ranges. Other dosage ranges can be $1x10^4$ to $1x10^{10}$ cells per kg/bodyweight. The total number of lymphocytes can be administered in a single bolus dose, or can be administered in two or more doses, such as one or more days apart.. The total number of lymphocytes can be administered in a single bolus dose, or can be administered in two or more doses, such as one or

more days apart. The amount of compound administered will depend on the potency of the specific lymphocyte composition, the disease being treated and the route of administration.

[00412] Additionally, the doses can comprise a mixture of lymphocytes, such as a mix of CD8+ and CD4+ cells. If a mix of CD8+ and CD4+ cells is provided, the ratio of CD8+ to CD4+ cells can be for example, 1:1, 1:2 or 2:1, 1:3 or 3:1, 1:4 or 4:1, 1:5 or 5:1, etc.

[00413] Modified lymphocytes can be administered prior to, concurrent with, or subsequent to other agents. If provided at the same time as other agents, the modified lymphocytes can be provided in the same or in a different composition. Thus, the lymphocytes and other agents can be presented to the individual by way of concurrent therapy. By "concurrent therapy" is intended administration to a subject such that the therapeutic effect of the combination of the substances is caused in the subject undergoing therapy. For example, concurrent therapy may be achieved by administering a dose of a pharmaceutical composition comprising modified lymphocytes and a dose of a pharmaceutical composition comprising at least one other agent, such as another chemotherapeutic agent, which in combination comprises a therapeutically effective dose, according to a particular dosing regimen. Similarly, modified lymphocytes and therapeutic agents can be administered in at least one therapeutic dose. Administration of the separate pharmaceutical compositions can be performed simultaneously or at different times (e.g., sequentially, in either order, on the same day, or on different days), as long as the therapeutic effect of the combination of these substances is caused in the subject undergoing therapy.

[00414] As described herein, the engineered Type I Cascade effector complexes of the present invention provide genome editing tools. Experiments demonstrating the functional reconstitution of Class 1 CRISPR-Cas system in mammalian cells for genome editing show that such streamlined plasmid designs can allow use of other Class 1 CRISPR-Cas systems, including those that exhibit fewer protein components and unique PAM requirements, and potentially even RNA- and DNA-targeting effector complexes from Type III CRISPR–Cas systems (see, e.g., Hille, F., et al., Cell 172:1239-1259 (2018); Tamulaitis, G., et al., Trends Microbiol. 25:49-61 (2017)). The multi-subunit nature of Cascade complexes provides the potential for multivalent and/or sterically precise recruitment of effector fusions such as synthetic transcription factors, epigenome modifiers, and base editors. In addition, heterologous expression of the complete DNA interference pathway from Type I systems – namely, Cascade-mediated recruitment of the Cas3 helicase-nuclease to genomic target sites – can be exploited to generate large DNA deletions, to expose long ssDNA tracts for homology-directed repair, and/or to mechanically disrupt protein-DNA roadblocks at defined genomic loci. Accordingly,

in one embodiment of the present invention, engineered Class 1 CRISPR-Cas systems can be used to generate large deletion regions and a donor polynucleotide (e.g., comprising suitable homology arms) can be introduced into the cell, thus facilitating insertion of at least a portion of the donor polynucleotide into the region.

[00415] Embodiments of the present invention include, but are not limited to, the following.

[00416] Embodiment 1. A composition comprising:

[00417] a first engineered Class 1 Type I CRISPR-Cas effector complex comprising,

[00418] a first Cse2 subunit protein, a first Cas5 subunit protein, a first Cas6 subunit protein, and a first Cas7 subunit protein,

[00419] a first fusion protein comprising a first Cas8 subunit protein and a first FokI, wherein the N-terminus of the first Cas8 subunit protein or the C-terminus of the first Cas8 subunit protein is covalently connected by a first linker polypeptide to the C-terminus or N-terminus, respectively, of the first FokI, and wherein the first linker polypeptide has a length of between about 10 amino acids and about 40 amino acids, and

[00420] a first guide polynucleotide comprising a first spacer capable of binding a first nucleic acid target sequence; and

[00421] a second engineered Class 1 Type I CRISPR-Cas effector complex comprising,

[00422] a second Cse2 subunit protein, a second Cas5 subunit protein, a second Cas6 subunit protein, and a second Cas7 subunit protein,

[00423] a second fusion protein comprising a second Cas8 subunit protein and a second FokI, wherein the N-terminus of the second Cas8 subunit protein or the C-terminus of the second Cas8 subunit protein is covalently connected by a second linker polypeptide to the C-terminus or N-terminus, respectively, of the second FokI, and wherein the second linker polypeptide has a length of between about 10 amino acids and about 40 amino acids, and

[00424] a second guide polynucleotide comprising a second spacer capable of binding a second nucleic acid target sequence, wherein a protospacer adjacent motif (PAM) of the second nucleic acid target sequence and a PAM of the first nucleic acid target sequence have an interspacer distance between about 20 bp and about 42 bp.

[00425] Embodiment 2. The composition of embodiment 1, wherein the first linker polypeptide has a length of between about 15 amino acids and about 30 amino acids.

[00426] Embodiment 3. The composition of embodiment 2, wherein the first linker polypeptide has a length of between about 17 amino acids and about 20 amino acids.

[00427] Embodiment 4. The composition of any one of embodiments 1-3, wherein the second linker polypeptide has a length of between about 15 amino acids and about 30 amino acids.

[00428] Embodiment 5. The composition of embodiment 4, wherein the second linker polypeptide has a length of between about 17 amino acids and about 20 amino acids.

[00429] Embodiment 6. The composition of any preceding embodiment, wherein the length of the first linker polypeptide and the second linker polypeptide are the same.

[00430] Embodiment 7. The composition of any preceding embodiment, wherein the second nucleic acid target sequence and the first nucleic acid target sequence each has an interspacer distance between about 22 bp and about 40 bp.

[00431] Embodiment 8. The composition of embodiment 7, wherein the second nucleic acid target sequence and the first nucleic acid target sequence each has an interspacer distance between about 26 bp and about 36 bp.

[00432] Embodiment 9. The composition of embodiment 8, wherein the second nucleic acid target sequence and the first nucleic acid target sequence each has an interspacer distance between about 29 bp and about 35 bp.

[00433] Embodiment 10. The composition of embodiment 9, wherein the second nucleic acid target sequence and the first nucleic acid target sequence each has an interspacer distance between about 30 bp and about 34 base bp.

[00434] Embodiment 11. The composition of any preceding embodiment, wherein the first FokI and the second FokI are monomeric subunits capable of associating to form a homodimer.

[00435] Embodiment 12. The composition of any one of embodiments 1-10, wherein the first FokI and the second FokI are distinct monomeric subunits capable of associating to form a heterodimer.

[00436] Embodiment 13. The composition of any preceding embodiment, wherein the N-terminus of the first Cas8 subunit protein is covalently connected by the first linker polypeptide to the C-terminus of the first FokI.

[00437] Embodiment 14. The composition of any one of embodiments 1-12, wherein the C-terminus of the first Cas8 subunit protein is covalently connected by a first linker polypeptide to the N-terminus of the first FokI.

[00438] Embodiment 15. The composition of any preceding embodiment, wherein the N-terminus of the second Cas8 subunit protein is covalently connected by the second linker polypeptide to the C-terminus of the second FokI.

[00439] Embodiment 16. The composition of any one of embodiments 1-14, wherein the C-terminus of the second Cas8 subunit protein is covalently connected by a second linker polypeptide to the N-terminus of the second FokI.

[00440] Embodiment 17. The composition of any preceding embodiment, wherein the first Cas8 subunit protein and the second Cas8 subunit protein each comprises identical amino acid sequences.

[00441] Embodiment 18. The composition of any preceding embodiment, wherein the first Cse2 subunit protein and the second Cse2 subunit protein each comprises identical amino acid sequences, the first Cas5 subunit protein and the second Cas5 subunit protein each comprises identical amino acid sequences, the first Cas6 subunit protein and the second Cas6 subunit protein each comprises identical amino acid sequences, and the first Cas7 subunit protein and the second Cas7 subunit protein each comprises identical amino acid sequences.

[00442] Embodiment 19. The composition of any preceding embodiment, wherein the first guide polynucleotide comprises RNA.

[00443] Embodiment 20. The composition of any preceding embodiment, wherein the second guide polynucleotide comprises RNA.

[00444] Embodiment 21. The composition of any preceding embodiment, wherein genomic DNA comprises the PAM of the second nucleic acid target sequence and the PAM of the first nucleic acid target sequence.

[00445] Embodiment 22. A cell comprising: the composition of any preceding embodiment.

[00446] Embodiment 23. The cell of embodiment 22, wherein genomic DNA of the cell comprises the PAM of the second nucleic acid target sequence and the PAM of the first nucleic acid target sequence.

[00447] Embodiment 24. The cell of embodiment 22 or 23, wherein the cell is a prokaryotic cell.

[00448] Embodiment 25. The cell of embodiment 22 or 23, wherein the cell is a eukaryotic cell.

[00449] Embodiment 26. One or more nucleic acid sequences encoding the first Cse2 subunit protein, the first Cas5 subunit protein, the first Cas6 subunit protein, the first Cas7 subunit protein, the first fusion protein, and the first guide polynucleotide of any one of embodiments 1-21.

[00450] Embodiment 27. One or more nucleic acid sequences encoding the second Cse2 subunit protein, the second Cas5 subunit protein, the second Cas6 subunit protein, the second

Cas7 subunit protein, the second fusion protein, and the second guide polynucleotide of any one of embodiments 1-21.

[00451] Embodiment 28. One or more expression cassettes comprising the one or more nucleic acid sequences of embodiment 26, embodiment 27, or embodiment 26 and embodiment 27.

[00452] Embodiment 29. One or more vectors comprising the one or more expression cassettes of embodiment 28.

[00453] Embodiment 30. A method of binding a polynucleotide comprising the first nucleic acid target sequence and the second nucleic acid target sequence, the method comprising:

[00454] providing the composition of any one of embodiments 1-21 for introduction into a cell or a biochemical reaction; and

[00455] introducing the composition into the cell or the biochemical reaction, thereby facilitating contact of the first engineered Class 1 Type I CRISPR-Cas effector complex with the first nucleic acid target sequence and contact of the second engineered Class 1 Type I CRISPR-Cas effector complex with the second nucleic acid target sequence, resulting in binding of the first engineered Class 1 Type I CRISPR-Cas effector complex with the first nucleic acid target sequence and binding of the second engineered Class 1 Type I CRISPR-Cas effector complex with the second nucleic acid target sequence in the polynucleotide.

[00456] Embodiment 31. The method of embodiment 30, wherein genomic DNA comprises the polynucleotide.

[00457] Embodiment 32. A method of cutting a polynucleotide comprising the first nucleic acid target sequence and the second nucleic acid target sequence, the method comprising:

[00458] providing the composition of any one of embodiments 1-21 for introduction into a cell or a biochemical reaction; and

[00459] introducing the composition into the cell or the biochemical reaction, thereby facilitating contact of the first engineered Class 1 Type I CRISPR-Cas effector complex with the first nucleic acid target sequence and contact of the engineered second Class 1 Type I CRISPR-Cas effector complex with the second nucleic acid target sequence, resulting in cutting of the first nucleic acid target sequence by the first engineered Class 1 Type I CRISPR-Cas effector complex and cutting of the second nucleic acid target sequence by the second engineered Class 1 Type I CRISPR-Cas effector complex.

[00460] Embodiment 33. The method of embodiment 32, wherein genomic DNA comprises the polynucleotide.

[00461] Embodiment 34. A kit comprising: the composition of any one of embodiments 1-21; and a buffer.

[00462] Embodiment 35. A kit comprising: the one or more nucleic acid sequences of embodiment 26, embodiment 27, or embodiment 26 and embodiment 27; and a buffer.

[00463] Embodiment 36. A composition comprising:

[00464] an engineered Class 1 Type I CRISPR-Cas effector complex comprising,

[00465] a Cse2 subunit protein, a Cas5 subunit protein, a Cas6 subunit protein, and a Cas7 subunit protein,

[00466] a first fusion protein comprising a Cas8 subunit protein and a first FokI, wherein the N-terminus of the first Cas8 subunit protein or the C-terminus of the first Cas8 subunit protein is covalently connected by a first linker polypeptide to the C-terminus or N-terminus, respectively, of the first FokI, and

[00467] a guide polynucleotide comprising a spacer capable of binding a nucleic acid target sequence; and

[00468] a second fusion protein comprising an engineered Class 1 Type I CRISPR-Cas3 fusion protein comprising a dCas3* protein and a second FokI, wherein the N-terminus of the dCas3* protein or the C-terminus of the dCas3* protein is covalently connected by a second linker polypeptide to the C-terminus or N-terminus, respectively, of the second FokI, and wherein the first linker polypeptide has a length of between about 10 amino acids and about 40 amino acids, effector complex comprising,

[00469] Embodiment 37. The composition of embodiment 36, wherein the first linker polypeptide has a length of between about 5 amino acids and about 40 amino acids.

[00470] Embodiment 38. The composition of embodiment 36, wherein the second linker polypeptide has a length of between about 5 amino acids and about 40 amino acids.

[00471] Embodiment 39. A cell comprising: the composition of any one of embodiments 36 to 38.

[00472] Embodiment 40. The cell of embodiment 39, wherein the cell is a prokaryotic cell.

[00473] Embodiment 41. The cell of embodiment 39, wherein the cell is a eukaryotic cell.

[00474] Embodiment 42. One or more nucleic acid sequences encoding the Cse2 subunit protein, the Cas5 subunit protein, the Cas6 subunit protein, the Cas7 subunit protein, the first fusion protein, and the guide polynucleotide of any one of embodiments 36 to 38.

[00475] Embodiment 43. One or more nucleic acid sequences encoding the second fusion protein of any one of embodiments 36 to 38.

[00476] Embodiment 44. One or more expression cassettes comprising the one or more nucleic acid sequences of embodiment 42, embodiment 43, or embodiment 42 and embodiment 43.

[00477] Embodiment 45. One or more vectors comprising the one or more expression cassettes of embodiment 44.

[00478] Embodiment 46. A method of binding a polynucleotide comprising the nucleic acid target sequence, the method comprising:

[00479] providing the composition of any one of embodiments 36 to 38 for introduction into a cell or a biochemical reaction; and

[00480] introducing the composition into the cell or the biochemical reaction, thereby facilitating contact of the engineered Class 1 Type I CRISPR-Cas effector complex with the nucleic acid target sequence and contact of the second fusion protein with the engineered Class 1 Type I CRISPR-Cas effector complex, resulting in binding of the engineered Class 1 Type I CRISPR-Cas effector complex and the second fusion protein to the nucleic acid target sequence in the polynucleotide.

[00481] Embodiment 47. The method of embodiment 46, wherein genomic DNA comprises the polynucleotide.

[00482] Embodiment 48. A method of cutting a polynucleotide comprising the nucleic acid target sequence, the method comprising:

[00483] providing the composition of any one of embodiments 36 to 38 for introduction into a cell or a biochemical reaction; and

[00484] introducing the composition into the cell or the biochemical reaction, thereby facilitating contact of the first engineered Class 1 Type I CRISPR-Cas effector complex with the first nucleic acid target sequence and contact of the engineered second Class 1 Type I CRISPR-Cas effector complex with the second nucleic acid target sequence,

[00485] introducing the composition into the cell or the biochemical reaction, thereby facilitating contact of the engineered Class 1 Type I CRISPR-Cas effector complex with the nucleic acid target sequence and contact of the second fusion protein with the engineered Class

1 Type I CRISPR-Cas effector complex, resulting in cutting of the nucleic acid target sequence by the engineered Class 1 Type I CRISPR-Cas effector complex and the second fusion protein.

[00486] Embodiment 49. The method of embodiment 48, wherein genomic DNA comprises the polynucleotide.

[00487] Embodiment 50. A kit comprising: the composition of any one of embodiments 36 to 38; and a buffer.

[00488] Embodiment 51. A kit comprising: the one or more nucleic acid sequences of embodiment 42, embodiment 43, or embodiment 42 and embodiment 43; and a buffer.

[00489] Embodiment 52. An engineered Type I CRISPR Cas3 mutant protein ("mCas3 protein") capable of reduced movement along DNA relative to a wild-type Type I CRISPR Cas3 protein ("wtCas3 protein"), the mCas3 protein comprising:

[00490] about 95% or higher sequence identity to the corresponding wtCas3 protein,

[00491] a nuclear localization signal is covalently connected at the amino terminus, carboxy terminus, or both the amino and carboxy termini, and

[00492] one or more mutations that down-modulates helicase activity, wherein the engineered Type I CRISPR Cas3 mutant protein retains nuclease activity;

[00493] wherein the DNA is double-stranded DNA (dsDNA) comprising a target region comprising a nucleic acid target sequence;

[00494] wherein when the wtCas3 protein is associated with a corresponding Cascade nucleoprotein complex ("Cascade NP complex/wtCas3 protein"), and the Cascade NP complex comprises a guide comprising a spacer complementary to the nucleic acid target sequence, binding of the Cascade NP complex/wtCas3 protein to the nucleic acid target sequence facilitates cleavage in the target region of the DNA, thereby resulting in a deletion ("wtCas3-deletion"); and

[00495] wherein the mCas3 protein when it is associated with the Cascade NP complex ("Cascade NP complex/mCas3 protein) and binds the nucleic acid target sequence facilitates cleavage in the target region of the DNA, thereby resulting in a shorter deletion relative to the wtCas3-deletion.

[00496] Embodiment 53. The mCas3 protein of embodiment 53, wherein the one or more mutations are substitutions of amino acids.

[00497] Embodiment 54. The mCas3 protein any preceding embodiment, wherein the one or more mutations are in either the RecA1 region or RecA2 region of the helicase domain.

[00498] Embodiment 55. The mCas3 protein of any preceding embodiment, wherein the one or more mutations down-modulate binding of the mCas3 protein to single-stranded DNA (ssDNA) relative to the wtCas3 protein.

[00499] Embodiment 56. The mCas3 protein of any preceding embodiment, wherein the one or more mutations down-modulate hydrolysis of adenosine triphosphate (ATP) by the mCas3 protein or down-modulate binding of ATP to the mCas3 protein relative to the wtCas3 protein.

[00500] Embodiment 57. The mCas3 protein of any preceding embodiment, wherein coding sequences for the mCas3 protein are covalently connected to the amino terminus or carboxy terminus of coding sequences of a Cas protein of the Cascade NP complex.

[00501] Embodiment 58. The mCas3 protein of any preceding embodiment, wherein the one or more mutations down-modulate binding of the mCas3 protein to single-stranded DNA (ssDNA) relative to the wtCas3 protein.

[00502] Embodiment 59. The mCas3 protein of any preceding embodiment, wherein coding sequences for the mCas3 protein are covalently connected to the amino terminus or carboxy terminus of coding sequences of a Cas protein of the Cascade RNP complex.

[00503] Embodiment 60. The mCas3 protein of any preceding embodiment wherein the Cas protein is selected from the group consisting of Cse2, Cas8 protein, Cas7 protein, Cas6 protein, and Cas5 protein.

[00504] Embodiment 61. The mCas3 protein of any preceding embodiment, wherein the wtCas3 protein is an *E. coli* Type 1 CRISPR Cas3 protein.

[00505] Embodiment 62. The mCas3 protein of embodiment 61, wherein the one or more mutations are selected from the group consisting of D452H, A602V, and D452H and A602V.

[00506] Embodiment 63. The mCas3 protein of any preceding embodiment, wherein the DNA is within a cell.

[00507] Embodiment 64. The mCas3 protein of embodiment 63, wherein the cell is a eukaryotic cell.

[00508] Embodiment 65. The mCas3 protein of embodiment 64, wherein the eukaryotic cell is a mammalian cell (e.g., human cell).

[00509] Embodiment 66. One or more polynucleotide encoding the mCas3 protein of any one of embodiments 52 to 65.

[00510] Embodiment 67. A plasmid comprising a polynucleotide sequence encoding the mCas3 protein of any one of embodiments 52 to 65 operably linked to regulatory sequences for expression in a mammalian cell.

[00511] Embodiment 68. One or more plasmid comprising a polynucleotide sequence encoding the mCas3 protein of any one of embodiments 52 to 65, and one or more polynucleotides encoding protein components of a corresponding Type I CRISPR Cascade operably linked to regulatory sequences for expression in a mammalian cell.

[00512] Embodiment 69. The one or more plasmids of embodiment 68, further comprising a plasmid encoding one or more guide polynucleotide operably linked to regulatory sequences for expression in a mammalian cell.

[00513] Embodiment 70. A Type I CRISPR Cascade nucleoprotein complex comprising the mCas3 protein of any one of embodiments 52 to 65.

[00514] Embodiment 71. The Type I CRISPR Cascade nucleoprotein complex of embodiment 70, wherein the nucleoprotein complex is an RNP.

[00515] Although preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. From the present Specification and the Examples, one skilled in the art can ascertain essential characteristics of this invention, and without departing from the spirit and scope thereof, can make changes, substitutions, variations, and modifications of the invention to adapt it to various usages and conditions. Such changes, substitutions, variations, and modifications are also intended to fall within the scope of the present disclosure.

Experimental

[00516] Aspects of the present invention are illustrated in the following Examples. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, concentrations, percent changes, and the like) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, temperature is in degrees Centigrade and pressure is at or near atmospheric. It should be understood that these Examples are given by way of illustration only and are not intended to limit the scope of the present invention.

[00517]

Example 1

In silico Design of Polynucleotides Encoding Cascade Components

[00518] This Example provides a description of the design of polynucleotide components encoding Cascade using gene, protein, and CRISPR sequences derived from a Type I-E CRISPR-Cas system.

[00519] Table 15 presents polynucleotide DNA sequences of genes encoding the five proteins of Cascade from Type I-E, specifically from *E. coli* strain K-12 MG1655, as well as the amino acid sequences of the resulting protein components. Genomic sequences were obtained from NCBI Reference Sequence NZ_CP014225.1. In Table 15, polynucleotide sequences were either amplified from *E. coli* gDNA or manufacturer-produced polynucleotides encoding Cascade protein components that were codon-optimized specifically for expression in *E. coli* and also for expression in human cells.

[00520]

Table 15
Cas Protein DNA and Amino Acid Sequences

Protein	Type of sequence	DNA coding sequence	Amino acid sequence
Cas8	genomic	SEQ ID NO:1	SEQ ID NO:16
Cse2	genomic	SEQ ID NO:2	SEQ ID NO:17
Cas7	genomic	SEQ ID NO:3	SEQ ID NO:18
Cas5	genomic	SEQ ID NO:4	SEQ ID NO:19
Cas6	genomic	SEQ ID NO:5	SEQ ID NO:20
Cas8	*E. coli* codon-optimized	SEQ ID NO:6	SEQ ID NO:16
Cse2	*E. coli* codon-optimized	SEQ ID NO:7	SEQ ID NO:17
Cas7	*E. coli* codon-optimized	SEQ ID NO:8	SEQ ID NO:18
Cas5	*E. coli* codon-optimized	SEQ ID NO:9	SEQ ID NO:19
Cas6	*E. coli* codon-optimized	SEQ ID NO:10	SEQ ID NO:20
Cas8	*H. sapiens* codon-optimized	SEQ ID NO:11	SEQ ID NO:16
Cse2	*H. sapiens* codon-optimized	SEQ ID NO:12	SEQ ID NO:17
Cas7	*H. sapiens* codon-optimized	SEQ ID NO:13	SEQ ID NO:18
Cas5	*H. sapiens* codon-optimized	SEQ ID NO:14	SEQ ID NO:19
Cas6	*H. sapiens* codon-optimized	SEQ ID NO:15	SEQ ID NO:20

[00521] In addition, several fusion proteins comprising Cascade proteins were designed. Table 16 presents polynucleotide DNA sequences of genes encoding Cascade protein fusion proteins, as well as the amino acid sequences of the resulting protein components. In most instances, fusion proteins described in Table 16 include short tri-amino acid linkers connecting the two polypeptide sequences within the fusion construct; this linker typically comprises glycine-glycine-serine (GGS) or glycine-serine-glycine (GSG). The exact tri-amino acid linker sequences used in each particular fusion protein can be found in the full-length amino acid sequence in Table 16.

[00522]

Table 16 Cascade Fusion Protein Sequences					
Cascade protein	**Heterologous polypeptide**	**Heterologous polypeptide fused to the N- or C-terminus of the Cascade protein**	**DNA coding sequence**	**Expression system for DNA coding sequence**	**Amino acid sequence**
Cse2	Strep-tag® II-HRV3C	N	SEQ ID NO:390	*E. coli*	SEQ ID NO:391
Cse2	His6–HRV3C	N	SEQ ID NO:392	*E. coli*	SEQ ID NO:393
Cse2	NLS	N	SEQ ID NO:394	Mammalian	SEQ ID NO:395
Cas5	NLS	N	SEQ ID NO:396	Mammalian	SEQ ID NO:397
Cas6	NLS	N	SEQ ID NO:398	*E. coli*	SEQ ID NO:399
Cas6	NLS-HA	N	SEQ ID NO:400	*E. coli*	SEQ ID NO:401
Cas6	NLS	N	SEQ ID NO:402	Mammalian	SEQ ID NO:403
Cas7	NLS	C	SEQ ID NO:404	*E. coli*	SEQ ID NO:405
Cas7	HA-NLS	C	SEQ ID NO:406	*E. coli*	SEQ ID NO:407
Cas7	NLS	N	SEQ ID NO:408	Mammalian	SEQ ID NO:409
Cas8	His6-MBP-TEV	N	SEQ ID NO:410	*E. coli*	SEQ ID NO:411
Cas8	His6-MBP-TEV-NLS-FokI-linker	N	SEQ ID NO:412	*E. coli*	SEQ ID NO:413
Cas8	NLS	N	SEQ ID NO:414	Mammalian	SEQ ID NO:415
Cas8	NLS-HA-FokI-linker	N	SEQ ID NO:416	Mammalian	SEQ ID NO:417

[00523] The His6 (hexahistidine; SEQ ID NO:418) and Strep-tag™ II (GE Healthcare Bio-Sciences, Pittsburgh, PA) (SEQ ID NO:419) peptide tags on the Cse2 protein, when co-expressed with other Cascade proteins, enable purification of the complex via either Nickel-nitriloacetic acid (Ni-NTA) resin or Strep-Tactin™ (IBA GMBH LLC, Göttingen, Germany) resin, respectively. The HRV3C (human rhinovirus 3C) protease recognition sequence (SEQ ID NO:420) is cleaved by an HRV3C protease and can be used to remove N-terminal fusions from

a protein of interest. The NLS (nuclear localization signal; SEQ ID NO:421 peptide tag on the Cas6, Cas7, and/or Cas8 proteins enables nuclear trafficking in eukaryotic systems. The HA (hemagglutinin; SEQ ID NO:422) peptide tag on the Cas6 or Cas7 proteins enables detection of heterologous protein expression by Western blotting with an anti-HA antibody. The MBP (maltose binding protein; SEQ ID NO:423) peptide fusion is a solubilization tag that facilitates purification of the Cas8 protein. The TEV (tobacco etch virus) protease recognition sequence (SEQ ID NO:424) is cleaved by TEV protease and can be used to remove N-terminal fusions from a protein of interest. The FokI nuclease domain comprises the *Sharkey* variant described by Guo, et al. (Guo, J., et al., J. Mol. Biol. 400:96-107 (2010)), two monomeric FokI subunits associate to form a homodimer, and catalyze double-stranded DNA cleavage upon homo-dimerization. A linker sequence (SEQ ID NO:425) is used to fuse the FokI nuclease domain to the Cas8 protein.

[00524] Additional linker sequences of varying length and amino acid composition have been designed that connect the FokI nuclease domain to the Cas8 protein. These amino acid sequences can be found in Table 17.

[00525]

Table 17
Amino Acid Linker Sequences

SEQ ID NO:	Linker length (amino acids)	Amino acid sequence
SEQ ID NO:426	5	GGGGS
SEQ ID NO:427	8	TGPGAAAR
SEQ ID NO:428	10	GGSGSSGGSG
SEQ ID NO:429	12	TGPGAAARAASG
SEQ ID NO:430	15	GGSGSSGGSGSSGGS
SEQ ID NO:431	16	SGSETPGTSESATPES
SEQ ID NO:432	20	SGSETPGTSESATPESGGS G
SEQ ID NO:433	30	SGSETPGTSESATPESGGS GSSGGSGSSGG

[00526] Table 18 contains the polynucleotide DNA sequence of four minimal CRISPR arrays that, when transcribed into precursor crRNA and processed by the RNA endonuclease protein of Cascade, generate mature crRNAs that function as the guide RNA to target complementary DNA sequences in biochemical assays and in cell culture gene editing experiments.

[00527] The minimal CRISPR array comprises two repeat sequences (underlined, lower case) flanking a spacer sequence, which represents the guide portion of the crRNA. RNA

processing by the Cascade endonuclease protein generates a crRNA with repeat sequences on both the 5' and 3' ends, flanking the guide sequence. The CRISPR array may also be expanded to include three repeat sequences (underlined) flanking two spacer sequences, which represent the guide portions of two distinct crRNAs by RNA processing by the endonuclease Cascade protein. The arrays can be further expanded to include additional spacer sequences, if desired.

[00528]

Table 18
CRISPR Array Sequences

SEQ ID NO:	Cell type	Target	Minimal CRISPR array sequence
SEQ ID NO:434	*E. coli*	Bacteriophage λ J3 target	gagttccccgcgccagcggggataaaccgCCAGTGAT AAGTGGAATGCCATGTGGGCTGTCgagttc cccgcgccagcggggataaaccg
SEQ ID NO:435	*E. coli*	Bacteriophage λ L3 target	gagttccccgcgccagcggggataaaccgAGTGGCAG ATATAGCCTGGTGGTTCAGGCGGCgagttc cccgcgccagcggggataaaccg
SEQ ID NO:436	*E. coli*	Bacteriophage λ L3/J3 targets	gagttccccgcgccagcggggataaaccgCCAGTGAT AAGTGGAATGCCATGTGGGCTGTCgagttc cccgcgccagcggggataaaccgAGTGGCAGATAT AGCCTGGTGGTTCAGGCGGCgagttccccgcg ccagcggggataaaccg
SEQ ID NO:437	*H. sapiens* cell	*TRAC* gene	gagttccccgcgccagcggggataaaccgGTTGATTTG CCTGCATTGGTGTTACACAGTCTgagttccc cgcgccagcggggataaaccgTAAGTTGTGTTCTT CTTTGCCTAGGCCTTCAGgagttccccgcgcca gcggggataaaccg

[00529]

Example 2

Design of Bacterial Expression Vectors for Production of Cascade Effector Complexes

[00530] This Example describes the design of bacterial expression vectors that encode the Cascade-associated proteins, as well as a minimal CRISPR array comprising the guide sequence as described in Example 1. The construction of Cascade subunit protein expression systems for use with plasmids encoding minimal CRISPR arrays is described.

[00531] A single-plasmid Cascade protein expression system was constructed to express the proteins of either a complex of Cascade in *E. coli*, known as the CasBCDE complex (which contains the Cse2, Cas7, Cas5, and Cas6 proteins, but not the Cas8 protein), or the entire functional Cascade complex in *E. coli*. The single plasmid system comprises either the *cse2–cas7–cas5–cas6* operon, or the entire *cas8–cse2–cas7–cas5–cas6* operon on a single expression plasmid. The Cas8 protein can be expressed from its own expression plasmid, for use in

biochemical experiments where it is mixed together with the CasBCDE complex to reconstitute Cascade.

[00532] A starting plasmid for expression vector construction was used (see Brouns, S., et al., Science 321:960–964 (2008)). The single-plasmid Cascade protein expression system comprising a Cas operon was assembled as follows. The coding sequences for the *cas* genes were arranged in the order *cse2–cas7–cas5–cas6* (CasBCDE complex or *cas8–cse2–cas7–cas5–cas6* (full Cascade complex), and were separated by sequences corresponding to the wild-type bacterial gene arrangement (see NCBI Reference Sequence NZ_CP014225.1).

[00533] In order to append a polynucleotide sequence encoding an affinity tag (His6 or Strep-tag® II, IBA GMBH LLC, Göttingen, Germany), the corresponding coding sequence was inserted at the junction of the 3' end of the *cas8* gene and the 5' end of the *cse2* gene; these two open reading frames overlap in the wild-type gDNA sequence.

[00534] In order to append polynucleotide sequences encoding N-terminal NLS and/or NLS-HA tags onto the 5' end of the *cas6* gene, additional spacing was introduced between the *cas6* and upstream *cas5* genes, because these open reading frames overlap in the wild-type gDNA sequence, such that the Shine-Dalgarno sequence for the *cas6* gene is within the 3' portion of the *cas5* gene. A new Shine-Dalgarno sequence was inserted upstream of the new NLS-Cas6 or NLS-HA-Cas6 open reading frames, to improve translational efficiency.

[00535] In order to append polynucleotide sequences encoding C-terminal NLS and/or HA-NLS tags onto the 3' end of the *cas7* gene, additional spacing was introduced between the *cas7* and downstream *cas5* genes, because these open reading frames are in close proximity in the wild-type gDNA sequence, such that the Shine-Dalgarno sequence for the *cas5* gene is within the 3' portion of the *cas7* gene. A new Shine-Dalgarno sequence was inserted downstream of the new Cas7-NLS or Cas7-HA-NLS open reading frames, to improve translational efficiency for the *cas5* gene.

[00536] In order to append polynucleotide sequences encoding N-terminal NLS-FokI-linker fusions to the Cas8 protein, the corresponding coding sequences were inserted at the 5' end of the *cas8* gene.

[00537] The *cse2–cas7–cas5–cas6* and *cas8–cse2–cas7–cas5–cas6* operons were cloned into the pCDF (MilliporeSigma, Hayward, CA) vector backbone, which confers spectinomycin resistance due to the presence of the *aadA* gene. Transcription of the operon is driven by a T7 promoter and is under control of the Lac operator; the vector also encodes the LacI repressor. A T7 terminator was cloned downstream of the *cse2–cas7–cas5–cas6* or *cas8–cse2–cas7–cas5–cas6* operon. The vector contains a CDF origin of replication.

[00538] For expression of Cas8 or FokI-Cas8 fusion proteins, the *cas8* gene was cloned into a pET (MilliporeSigma, Hayward, CA) family vector backbone, which confers kanamycin resistance due to the presence of the *kanR* gene. Transcription of the operon is driven by a T7 promoter (P_{T7}), and is under control of the Lac operator (*lacO*); the vector also encodes the LacI repressor (*lacI* gene). A T7 terminator was cloned downstream of the *cas8* gene. The vector contains a ColE1 origin of replication.

[00539] FIG. 23A, FIG. 23B, FIG. 23C, FIG. 23D, and FIG. 23E present schematic diagrams of overexpression vectors for the *cas8*, *fokI–cas8*, the *cse2–cas7–cas5–cas6* operon, the *cas8–cse2–cas7–cas5–cas6* operon, and the *fokI-cas8–cse2–cas7–cas5–cas6* operon. The designations in FIG. 23A, FIG. 23B, FIG. 23C, FIG. 23D, and FIG. 23E are described in this Example (as well as in Example 1) and are as follows: P_{T7} (T7 promoter), *lacO* (Lac operator), His6 (hexahistidine), MBP (maltose binding protein), Strep-tag® II (IBA GMBH LLC, Göttingen, Germany) HRV3C (human rhinovirus 3C) protease recognition sequence, TEV (tobacco etch virus) protease recognition sequence, NLS (nuclear localization signal), *kanR* (kanamycin resistance gene), *lacI* (LacI repressor gene), colE1 ori (origin of replication), CDF ori (CloDF13 origin of replication), FokI nuclease domain (*Sharkey* variant), and *aadA* (gene encoding aminoglycoside resistance protein).

[00540] Table 19 provides sequences of bacterial expression plasmids encoding the Cas8 protein, the four proteins of the CasBCDE complex (*cse2–cas7–cas5–cas6* operon), and all five proteins of the Cascade complex (*cas8–cse2–cas7–cas5–cas6* operon). Polynucleotide sequences are provided with and without the N-terminal FokI fusion on the Cas8 protein.

[00541]

Table 19
Bacterial Plasmid Sequences

SEQ ID NO:	Vector designation	Arrangement of protein coding sequences (N to C)	Notable characteristics
SEQ ID NO:438	Cas8 expression vector	His6-MBP-TEV-Cas8	Can be added to CasBCDE complex to reconstitute Cascade
SEQ ID NO:439	FokI-Cas8 expression vector	His6-MBP-TEV-NLS-Fok1-linker-Cas8	FokI confers the ability to cleave double-stranded DNA
SEQ ID NO:440	CasBCDE complex expression vector	Strep-tag® II-HRV3C-Cse2_Cas7_Cas5_Cas6	When co-expressed with a CRISPR array, generates CasBCDE complex

Table 19 Bacterial Plasmid Sequences			
SEQ ID NO:	**Vector designation**	**Arrangement of protein coding sequences (N to C)**	**Notable characteristics**
SEQ ID NO:441	Cascade complex expression vector	Cas8_His6–HRV3C–Cse2_Cas7_Cas5_Cas6	When co-expressed with a CRISPR array, generates Cascade complex
SEQ ID NO:442	FokI-Cascade expression vector	NLS-FokI-linker-Cas8_His6–HRV3C–Cse2_Cas7_Cas5_Cas6	FokI confers the ability to cleave double-stranded DNA targeted by crRNA
SEQ ID NO:443	FokI-Cascade expression vector, extra NLS tag	NLS-FokI-linker-Cas8_His6–HRV3C–Cse2_Cas7-NLS_Cas5_Cas6	FokI confers the ability to cleave double-stranded DNA targeted by crRNA; extra NLS tag on Cas7 protein improves nuclear trafficking

[00542] In order to purify the CasBCDE complex and Cascade complex containing a crRNA, the protein expression vectors encoding the *cse2–cas7–cas5–cas6* operon or the *cas8–cse2–cas7–cas5–cas6* operon are combined with a vector containing a minimal CRISPR array.

[00543] CRISPR arrays were cloned into the pACYC-Duet1 vector backbone, which confers chloramphenicol resistance due to the *camR* gene. Transcription of the array is driven by a T7 promoter and is under control of the Lac operator (*lacO*); the vector also encodes the LacI repressor. A T7 terminator was cloned downstream of the CRISPR array. The vector contains a p15A origin of replication.

[00544] FIG. 24 contains a schematic diagram of an expression vector containing a CRISPR array with 2 repeats (FIG. 24, "repeats") and 1 spacer (FIG. 24, "spacer"). The array can be expanded, as described herein. The designations in FIG. 24 are described in this Example (as well as in Example 1) and are as follows: P_{T7} (T7 promoter), *lacO* (Lac operator), *lacI* (LacI repressor gene), p15A ori (origin of replication), and *camR* (chloramphenicol resistance gene).

[00545] Table 20 provides the sequences of bacterial expression plasmids encoding examples of minimal CRISPR arrays.

[00546]

Table 20 Bacterial Plasmid Sequences			
SEQ ID NO:	**Vector designation**	**DNA targeted by spacer**	**Notable characteristics**
SEQ ID NO:444	CRISPR(J3) expression vector	Bacteriophage λ J3 target	Two repeats, one spacer
SEQ ID NO:445	CRISPR(L3) expression vector	Bacteriophage λ L3 target	Two repeats, one spacer
SEQ ID NO:446	CRISPR(J3/L3) expression vector	Bacteriophage λ L3/J3 targets	Three repeats, two spacers
SEQ ID NO:447	CRISPR(*TRAC*) expression vector	*TRAC* gene	Three repeats, two spacers

[00547]

Example 3

Design of Eukaryotic Expression Vectors for Production of Cascade Effector Complexes in Mammalian Cells

[00548] This Example describes the design of eukaryotic expression plasmid vectors that encode Cascade-associated proteins, as well as minimal CRISPR arrays comprising the component sequences as described in Example 1.

[00549] A. Separate plasmids expressing each Cascade protein and minimal CRISPR array

[00550] Cascade proteins can be expressed in mammalian cells by encoding each of the protein components on a separate expression vector driven by the human cytomegalovirus (CMV) immediate-early promoter/enhancer and encoding the crRNA on a separate expression vector driven by the human U6 promoter.

[00551] The starting plasmid for each expression plasmid was a derivative of pcDNA3.1 (Thermo Scientific, Wilmington, DE). Coding sequences for the Cascade proteins, codon-optimized for expression in human cells (see Example 1), were inserted into the vector downstream of the CMV promoter and upstream of a bovine growth hormone (bGH) polyadenylation signal. The *cse2* gene was fused to polynucleotide sequences at the 5' end coding for an N-terminal NLS and 3x-FLAG epitope tag. The *cas5* gene was fused to polynucleotide sequences at the 5' end coding for an N-terminal NLS. The *cas6* gene was fused to polynucleotide sequences at the 5' end coding for an N-terminal NLS and HA epitope tag. The *cas7* gene was fused to polynucleotide sequences at the 5' end coding for an N-terminal NLS and Myc epitope tag. The *cas8* gene was fused to polynucleotide sequences at the 5' end

coding for an N-terminal NLS; in another embodiment, the *cas8* gene was fused to polynucleotide sequences at the 5' end coding for an N-terminal NLS, HA epitope tag, and FokI nuclease domain.

[00552] Each gene or gene fusion was cloned into a pcDNA3.1 derivative vector backbone, which confers ampicillin resistance due to the presence of the *ampR* gene. The vector also encodes neomycin resistance due to the presence of the *neoR* gene, which is downstream of an SV40 early promoter (P_{SV40}) and origin (SV40 ori), and upstream of an SV40 early polyadenylation signal (SV40 pA). In addition to the human CMV immediate-early promoter/enhancer (P_{CMV}) and bGH (bovine growth hormone) polyadenylation signal, the vector contains a T7 promoter upstream of the gene of interest, allowing for *in vitro* transcription of mRNA. The vector contains an f1 origin of replication as well as a ColE1 origin of replication.

[00553] FIG. 25 contains a schematic diagram of a mammalian expression vector encoding the FokI-Cas8 fusion protein. The designations in FIG. 25 are described in this Example (as well as in Example 1) and are as follows: the human CMV immediate-early promoter/enhancer (P_{CMV}), NLS (nuclear localization signal), FokI (FokI nuclease domain (*Sharkey* variant)), Cas8 protein coding sequence, bGH pA (bovine growth hormone polyadenylation signal), f1 ori (f1 phage origin of replication), P_{SV40} (SV40 early promoter), SV40 ori (SV40 origin), *neoR* (neomycin resistance gene), SV40 pA (SV40 early polyadenylation signal), colE1 ori (origin of replication), and *ampR* (ampicillin resistance gene). Vectors encoding the other Cascade proteins were designed similarly.

[00554] Table 21 provides the sequences of individual mammalian expression vectors encoding each of Cse2, Cas5, Cas6, Cas7, Cas8, and FokI-Cas8.

[00555]

Table 21
Mammalian Expression Vectors

SEQ ID NO:	**Vector designation**	**Notable characteristics**
SEQ ID NO:448	Mammalian Cse2 expression vector	Cse2 contains N-terminal NLS-3xFLAG tag
SEQ ID NO:449	Mammalian Cas5 expression vector	Cas5 contains N-terminal NLS
SEQ ID NO:450	Mammalian Cas6 expression vector	Cas6 contains N-terminal NLS-Ha tag
SEQ ID NO:451	Mammalian Cas7 expression vector	Cas7 contains N-terminal NLS-Myc tag
SEQ ID NO:452	Mammalian Cas8 expression vector	Cas8 contains N-terminal NLS

Table 21 Mammalian Expression Vectors		
SEQ ID NO:	**Vector designation**	**Notable characteristics**
SEQ ID NO:453	Mammalian FokI-Cas8 expression vector	Cas8 contains N-terminal NLS-HA-FokI; FokI confers the ability to cleave double-stranded DNA

[00556] The CRISPR RNA was encoded with a minimal CRISPR array containing three repeats flanking two spacer sequences. The construct generating CRISPR RNA can be designed with additional sequences flanking the outermost repeats in the minimal array. Processing of the precursor CRISPR RNA is enabled by the RNA processing protein of the Cascade complex (Cas6 protein), which can be expressed on a separate plasmid.

[00557] The CRISPR array was cloned into the same pcDNA3.1 derivative vector backbone described above, except the human CMV promoter was replaced with the human U6 promoter (P_{U6}), and the bGH polyadenylation signal was replaced with a poly-T termination signal. An example of such a CRISPR array is illustrated in FIG. 35. In the figure, the hU6 promoter (FIG. 35, shown as a stippled region) is adjacent a first repeat sequence (clear square), that is adjacent a first spacer sequence (FIG. 35, spacer 1, slanted lines), that is adjacent a second repeat sequence (FIG. 35, grey square), that is adjacent a second spacer sequence (FIG. 35, spacer 2), that is adjacent a third repeat sequence (FIG. 35, black square). In FIG. 35, the region comprising the paired gRNA guides is shown (FIG. 35, paired gRNAs).

[00558] FIG. 26 contains a schematic diagram of a eukaryotic expression vector encoding a representative CRISPR array targeting the *TRAC* gene. The designations in FIG. 26 are described in this Example (as well as in Example 1) and are as follows: P_{U6} (human U6 promoter), repeats (CRISPR RNA repeats), *TRAC* spacer-1 (first spacer targeting the *TRAC* gene), *TRAC* spacer-2 (second spacer targeting the *TRAC* gene), polyT (poly-T termination signal), f1 ori (f1 phage origin of replication), P_{SV40} (SV40 early promoter), SV40 ori (SV40 origin), *neoR* (neomycin resistance gene), SV40 pA (SV40 early polyadenylation signal), colE1 ori (origin of replication), and *ampR* (ampicillin resistance gene).

[00559] Table 22 provides the sequence of a representative mammalian expression vector encoding a CRISPR array targeting the *TRAC* gene; a spacer sequence that targets matching DNA sequences in the *TRAC* gene can be found in Table 18.

[00560]

Table 22 Mammalian Expression Vector			
SEQ ID NO:	**Vector designation**	**Spacer complementary to target**	**Notable characteristics**
SEQ ID NO:454	Mammalian CRISPR RNA expression vector	*TRAC* gene	Three repeats, two spacers

[00561] B. Cascade protein expression system wherein multiple Cascade protein coding sequences are expressed from a single promoter

[00562] In order to express components of the Cascade complex from fewer expression vectors, polycistronic expression vectors were constructed. On each, a single CMV promoter drives expression of multiple coding sequences simultaneously that are separated by a 2A viral peptide sequence. The *Thosea asigna* virus 2A peptide sequence induces ribosomal skipping (see, e.g., Liu, Z., et al., Sci. Rep. 7:2193 (2017)), thus enabling multiple protein-coding genes to be concatenated within a single polycistronic construct.

[00563] The starting plasmid for the polycistronic expression plasmid was the same derivative of pcDNA3.1 described above, containing the CMV promoter and bGH polyadenylation signal. Coding sequences for the Cascade proteins, codon-optimized for expression in human cells (see Example 1), were joined in the order *cas7-cse2-cas5-cas6-cas8*, with a polynucleotide sequence coding for the *Thosea asigna* virus 2A (T2A) peptide inserted in between each pair of genes. In addition, polynucleotide sequences encoding NLS tags were appended to the 5' end of each Cascade protein gene, and a polynucleotide sequence encoding the FokI nuclease domain was appended to the 5' end of the *cas8* gene, connecting by a 30-amino acid linker sequence. The final construct has the following order of elements: *NLS-cas7-T2A-NLS-cse2-T2A-NLS-cas5-T2A-NLS-cas6-T2A-NLS-fokI-linker-cas8*.

[00564] FIG. 27 contains a schematic diagram of an exemplary polycistronic mammalian expression vector encoding all the Cascade proteins. The designations in FIG. 27 are described in this Example (as well as in Example 1) and are as follows: the human CMV immediate-early promoter/enhancer (P_{CMV}), NLS (nuclear localization signal), T2A (polynucleotide sequence coding for the *Thosea asigna* virus 2A peptide), coding sequences for the Cas7, Cse2, Cas5, and Cas6 proteins, fokI (FokI nuclease domain (*Sharkey* variant) a linker sequence, coding sequence for Cas8 protein, bGH pA (bovine growth hormone polyadenylation signal), f1 ori (f1 phage origin of replication), P_{SV40} (SV40 early promoter), SV40 ori (SV40 origin), *neoR* (neomycin resistance gene), SV40 pA (SV40 early polyadenylation signal), colE1 ori (origin of replication), *ampR* (ampicillin resistance gene), and an MluI restriction site.

[00565] Table 23 provides the sequence of an exemplary polycistronic mammalian expression vector encoding all the Cascade proteins. This vector can be combined with the mammalian expression vector encoding CRISPR RNA described above to produce functional Cascade complexes in mammalian cells.

[00566]

Table 23 **Mammalian Expression Vectors**			
SEQ ID NO:	**Vector designation**	**Arrangement of protein coding sequences (N to C)**	**Notable characteristics**
SEQ ID NO:455	Polycistronic mammalian expression vector encoding all five Cascade proteins	NLS-Cas7-T2A_NLS-Cse2-T2A_NLS-Cas5-T2A_NLS-Cas6-T2A_NLS-FokI-Cas8	Single protein expression vector encoding all Cascade proteins, each with N-terminal NLS tag. Cas8 contains N-terminal NLS-HA-FokI; FokI confers the ability to cleave double-stranded DNA

[00567] C. Single plasmid expression system

[00568] A single-plasmid Cascade expression system was constructed to express the complete Cascade complex in human cells. The plasmid encodes the entire *cas8–cse2–cas7–cas5-cas6* operon and a minimal CRISPR array on a single plasmid. This plasmid was constructed from the polycistronic protein expression vector (described in Table 23 and FIG. 27) by inserting the minimal CRISPR array along with the upstream human U6 promoter and downstream poly-T termination signal into the MluI restriction site.

[00569] Table 24 provides the sequence of the single plasmid for expression of all five Cascade proteins together with the crRNA to facilitate formation of Cascade complexes in human cells.

[00570]

Table 24 Mammalian Expression Vector			
SEQ ID NO:	**Vector designation**	**Arrangement of protein coding sequences (N to C)**	**Notable characteristics**
SEQ ID NO:456	Polycistronic mammalian expression vector encoding all five Cascade proteins and crRNA	hU6_CRISPR(*TRAC*), CMV_NLS–Cas7-T2A_NLS–Cse2-T2A_NLS–Cas5-T2A_NLS–Cas6_NLS–FokI–Cas8	Single protein expression vector encoding crRNA and all Cascade proteins, each with N-terminal NLS tag. Cas8 contains N-terminal NLS-HA-FokI; FokI confers the ability to cleave double-stranded DNA

[00571] Plasmids were also designed for the expression of the Cas3 protein (SEQ ID NO:21; monomer Cas3 nuclease/helicase *E. coli* K-12 substr. MG1655) in *E. coli* and in mammalian cells. Table 25 provides the constructs and sequences of these plasmids.

[00572]

Table 25 Cas3 Protein Fusions		
SEQ ID NO:	**Protein**	**Notable characteristics**
SEQ ID NO:457	Cas3	gDNA gene sequence
SEQ ID NO:458	Cas3	Protein amino acid sequence
SEQ ID NO:459	His6-MBP-TEV-Cas3	Derived from gDNA gene sequence
SEQ ID NO:460	His6-MBP-TEV-Cas3	Protein amino acid sequence
SEQ ID NO:461	His6-MBP-TEV-Cas3	Cas3 *E. coli* expression vector
SEQ ID NO:462	Cas3, human codon-optimized	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:463	Cas3-NLS	*H. sapiens* codon-optimized DNA gene sequence
SEQ ID NO:464	Cas3-NLS	Protein amino acid sequence
SEQ ID NO:465	Cas3-NLS	Cas3 mammalian expression vector

[00573]

Example 4

Introduction of Polynucleotides Encoding Cascade Components into a Bacterial Production Strain

[00574] This Example describes for introduction and expression of Cas8 subunit protein coding sequences, as well as coding sequences for components of engineered Type I CRISPR-Cas effector complexes in bacterial cells using *E. coli* expression systems.

[00575] A. Expression of Cas8 protein

[00576] *E. coli* Type I-E Cas8 protein was expressed from a plasmid (Example 2, SEQ ID NO:438, Table 19, FIG. 23A) containing an operon for the IPTG inducible expression of His6-MBP-TEV-Cas8 from a T7 promoter. The expression plasmid conferred resistance to kanamycin.

[00577] In order to express Cas8 protein, *E. coli* cells were transformed with the expression plasmid. Briefly, a 100 µL aliquot of chemically competent *E. coli* cells (*E. coli* BL21 Star™ (Thermo Fisher Scientific, Waltham, MA) cells) in a microcentrifuge tube was thawed on ice for 10 minutes. 35 ng of plasmid DNA was added to the thawed cells and the cells were incubated with the DNA on ice for 8 minutes. Heat shock was performed by a placing the microcentrifuge tube in a 42°C water bath for 30 seconds and then immediately placing the tube in ice for 2 minutes. 900 µL of 2xYT media were added to the microcentrifuge tube, and the microcentrifuge tube was placed in a tube rotator at 37°C for 1 hour. Finally, 100 µL of the recovered cells were plated on LB solid kanamycin (50 µg/mL) and incubated overnight at 37°C.

[00578] A single colony was picked from the colonies that grew on the antibiotic selection plates and was inoculated into 10 mL of 2xYT media supplemented with kanamycin (50 µg/mL). The culture was grown overnight at 37°C while shaking in an orbital shaker at 200 RPMs. 6 mL of the overnight culture were transferred to a 2 L baffled flask having 1 L of 2xYT media supplemented with kanamycin (50 µg/mL). The 1 L culture was grown at 37°C while shaking in an orbital shaker at 200 RPM until the optical density at 600 nm was 0.56.

[00579] Expression was then induced by the addition of IPTG to a final concentration of 1 mM. The induced cultures were grown overnight at 16°C while shaking in an orbital shaker at 200 RPM. Cells were harvested by centrifugation at 4,000 RCF for 15 minutes at 4°C. The cell pellet was re-suspended in 15 mL of a lysis buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, and 1 mM TCEP supplemented with 1 Complete™ (Roche, Basel, Switzerland) protease inhibitor tablet per 50 mL of lysis buffer. The re-suspended cells were transferred to a 50 mL conical tube for immediate downstream processing. The Cas8 protein was purified and the purified protein characterized essentially as described below for the FokI-Cas8 fusion protein (Example 5C).

[00580] B. Expression of the components of Cascade RNP complexes

[00581] A complete set of the five *E. coli* Cascade proteins and RNA guides were co-expressed in *E. coli* cells using a two-plasmid system to produce Cascade RNP complexes. One plasmid (Example 2, SEQ ID NO:441, Table 19, FIG. 23D) contained an operon for IPTG inducible expression of the Cse2, Cas5, Cas6, Cas7, and Cas8 proteins from a T7 promoter. A

His6 affinity tag was included as a translational fusion to the N-terminus of Cse2 (Example 1, SEQ ID NO:392, Table 16). The second plasmid coded for the IPTG inducible expression of the J3 guide (Example 2, SEQ ID NO:444, Table 20, FIG. 24). The Cascade protein expression plasmid conferred spectinomycin resistance, and the Cascade RNA guide expression plasmid conferred chloramphenicol resistance.

[00582] In order to co-express the Cascade proteins and RNA components in the same cell, *E. coli* cells were simultaneously transformed with the two plasmids. A 100 µL aliquot of chemically competent *E. coli* cells (*E. coli*, BL21 Star™ (DE3) (Thermo Fisher Scientific, Waltham, MA)) in a microcentrifuge tube was thawed on ice for 10 minutes. 35 ng of each plasmid was added to the thawed cells and the cells were incubated with the DNA on ice for 8 minutes. Heat shock was performed by a placing the microcentrifuge tube in a 42°C water bath for 30 seconds and then immediately placing the microcentrifuge tube in ice for 2 minutes. 900 µL of 2xYT media were added to the microcentrifuge tube and the microcentrifuge tube placed in a tube rotator at 37°C for 1 hour. Finally, 100 µL of the recovered cells were plated on LB solid media with chloramphenicol (34 µg/mL) and spectinomycin (50 µg/mL) and incubated overnight at 37°C.

[00583] A single colony was picked from the colonies that grew on the antibiotic selection plates and was inoculated into 10 mL of 2xYT media supplemented with chloramphenicol (34 µg/mL) and spectinomycin (100 µg/mL). The culture was grown overnight at 37°C while shaking in an orbital shaker at 200 RPMs. 6 mL of the overnight culture were transferred to a 2 L baffled flask having 1 L of 2xYT media supplemented with chloramphenicol (34 µg/mL) and spectinomycin (100 µg/mL). The 1 L culture was grown at 37°C while shaking in an orbital shaker at 200 RPM until the optical density at 600 nm was 0.56.

[00584] Expression from both plasmids was induced by the addition of IPTG to a final concentration of 1 mM. The induced cultures were grown overnight at 16°C while shaking in an orbital shaker at 200 RPM. Cells were harvested by centrifugation for at 4,000 RCF for 15 minutes at 4°C. The cell pellet was re-suspended in 15 mL of lysis buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, and 1 mM TCEP supplemented with 1 Complete™ (Roche, Basel, Switzerland) protease inhibitor tablet per 50 mL of lysis buffer. The re-suspended cells were transferred to a 50 mL conical tube for immediate downstream processing. Cascade RNP complexes were purified and characterized as described below.

[00585]

Example 5

Purification of Cascade Components and Cascade RNP Complexes

[00586] This Example describes a method to purify *E. coli* Type I-E Cascade RNP complexes produced by overexpression in bacteria as described in Example 4B. The method uses immobilized metal affinity chromatography followed by size exclusion chromatography (SEC). This Example also describes the methods used to assess the quality of the purified Cascade RNP product. In addition, this Example describes purification and characterization of Cascade components.

[00587] A. Purification of Cas8, Cas7, Cas6, Cas5, and Cse2 Cascade RNP complexes

[00588] *E. coli* Type I-E Cascade RNP complexes were produced as described in Example 4B. The Cascade complexes were captured using immobilized metal affinity chromatography. Briefly, the re-suspended cell pellets, produced as described in Example 4B, were thawed on ice and the volume was brought to 35 mL by of an additional 15 mL of lysis buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, and 1 mM TCEP supplemented with 1 Complete™ (Roche, Basel, Switzerland) protease inhibitor tablet per 50 mL of lysis buffer.

[00589] The 50 mL conical tube was placed in an ice water bath and the cells were lysed by two rounds of sonication using a Q500 sonicator with a ½ inch tip (Qsonica, Newtown, CT). Each round of sonication consisted of a treatment cycle of 2.5 minutes with repeating cycles of 10 seconds of sonication at 50% amplitude followed by 20 seconds of rest. The tube was allowed to cool in the ice water bath for one minute between rounds of sonication. The lysates were clarified by centrifugation at 48,384 RCF for 30 minutes at 4°C. The clarified supernatant was then added to a Hispur™ Ni-NTA (Thermo Fisher Scientific, Waltham, MA) resin, that had been pre-equilibrated with Ni-wash buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 10 mM imidazole, 5% glycerol, and 1 mM TCEP. A 1.5 mL bed volume of nickel affinity resin was used for each 1 L of *E. coli* expression culture. After one hour of incubation at 4°C with gentle mixing, the resin was pelleted by centrifugation at 500 RCF for 2 minutes at 4°C. The supernatant was aspirated and the resin was washed 5 times with 5 bed volumes of Ni-wash buffer. After each wash the resin was pelleted at 500 RCF for 2 minutes at 4°C and the supernatant was removed by aspiration. Finally, bound proteins (including the Cascade RNP complexes) were eluted by the addition of five bed volumes of Ni-elution buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 300 mM imidazole, 5% glycerol, and 1 mM Tris(2-

carboxyethyl)phosphine (TCEP). After centrifugation at 500 RCF for 2 minutes at 4°C, the nickel affinity eluate was aspirated into a clean 50 mL conical tube.

[00590] The nickel affinity eluate was further purified by size exclusion chromatography (SEC). The nickel affinity eluate was concentrated to a final volume of 0.5 mL by ultrafiltration at 12°C using an Amicon® (MilliporeSigma, Billerica, MA) ultrafiltration spin concentrator with an Ultracel®-50 (MilliporeSigma, Hayward, CA) membrane. The concentrated sample was filtered using a 0.22 μM Ultrafree-MC GV (MilliporeSigma, Hayward, CA) centrifugal filter before being further purified by separation at 4°C with a flow rate of 0.5 mL/minute on a HiPrep™ 16/60 Sephacryl® S-300 (GE Healthcare, Uppsala, Sweden) column equilibrated with SEC buffer composed of 50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 0.1 mM EDTA, and 1 mM TCEP. Proteins were eluted with SEC buffer and 1 ml fractions were collected. The earliest eluting peak, as judged by UV 280, was assumed to be high molecular weight aggregated material and the corresponding fractions were discarded. Subsequent elution fractions were analyzed by Coomassie stained SDS-PAGE. Each properly formed complex contained one molecule of Cas8, six molecules of Cas7, one molecule each of Cas6 and Cas5, and two molecules of Cse2. Elution fractions that had the approximate expected stoichiometry of Cascade proteins, when visualized on the SDS-PAGE gel, were pooled. Pooled fractions were analyzed spectrophotometrically to confirm they contained a significant nucleic acid component, as evidenced by an absorbance at 260 nm that is greater than the absorbance at 280 nm.

[00591] The pooled samples were exchanged into storage buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, 0.1 mM EDTA, and 1 mM TCEP by concentrating the pooled samples to 100 uL with an Amicon® (MilliporeSigma, Hayward, CA) spin concentrator with an Ultracel®-50 (MilliporeSigma, Hayward, CA) membrane and then diluting 50-fold with the storage buffer. Finally, the sample was concentrated to 10 mg/mL using the same ultrafiltration device and stored at -80°C.

[00592] The final purified product was analyzed spectrophotometrically to determine the final concentration of the Cascade RNP complexes and to confirm the presence of a nucleic acid component as evidenced by an absorbance at 260 nm that is greater than the absorbance at 280 nM. The concentration of the Cascade RNP complexes was determined by dividing the absorbance at 280 nm by the calculated absorbance of a 0.1% solution of the intact complex with a 1 cm path length. The predicted absorbance of a 0.1% solution of the purified complex is 2.03 cm^{-1} and was calculated by dividing the sum of the calculated extinction coefficients at

280 nm for each of the molecules in the complex (916940 $M^{-1}cm^{-1}$) by the sum of the molecular weights of each of the molecules in the complex (450832 g/mole).

[00593] Additionally, the final product was analyzed by SDS-PAGE with Coomassie blue staining to confirm that each protein component was present in approximately the correct stoichiometry, and to assess the presence of contaminating proteins. SDS-PAGE gels were stained with a Coomassie InstantBlue™ (Expedeon, San Diego, CA) stain. Gels were imaged using a Gel doc™ EZ (Bio-Rad, Hercules, CA) imager and annotated using ImageLab (Bio-Rad, Hercules, CA) software.

[00594] B. Purification of Cascade complexes comprising Cas7, Cas6, Cas5, and Cse2 proteins

[00595] A Cascade complex composed of the and the protein components Cas7, Cas6, Cas5, and Cse2 was purified. The L3 guide RNA (Example 2, SEQ ID NO:445, Table 20) was expressed from a first plasmid (Example 2, FIG. 24) essentially as described in Example 4B. The Cascade proteins were expressed from a second plasmid (Example 2, SEQ ID NO:440, Table 19, FIG. 23C) essentially as described in Example 4B.

[00596] The complex was captured using affinity chromatography. Re-suspended cell pellets were thawed on ice. In a 50 mL conical tube, the volume was brought up to 35 mL with an additional 15 mL of lysis buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, 1 mM TCEP, and supplemented with 1 Complete™ (Roche, Basel, Switzerland) protease inhibitor tablet per 50 mL of lysis buffer. The 50 mL conical tube was placed in an ice water bath, and the cells were lysed by six rounds of sonication using a Q500 sonicator with a ½ inch tip (Qsonica, Newtown, CT). Each round of sonication consisted of a 1-minute treatment cycle with repeating cycles of 3 seconds of sonication at 90% amplitude followed by 9 seconds of rest. The tube was allowed to cool in the ice water bath for one minute between rounds of sonication. The lysate was clarified by centrifugation at 48,384 RCF for 30 minutes at 4°C. The clarified supernatant was affinity purified by addition of Strep-Tactin® Sepharose® (IBA GMBH LLC, Göttingen, Germany) resin that had been pre-equilibrated with Strep-wash buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 5% glycerol, and 1 mM TCEP. A 0.55 mL bed volume of affinity resin was used for each 1 L of *E. coli* expression culture. After one hour of incubation at 4°C with gentle mixing, the sample was poured onto a 30 mL disposable gravity flow column (Bio-Rad, Hercules, CA) allowing the unbound material to flow through the column. The resin was washed five times with five bed volumes of Strep-wash buffer. Finally, the bound proteins were eluted with two sequential additions of five bed

volumes of Strep-elution buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 2.5mM Desthiobiotin, 5% glycerol, 1 mM EDTA, and 1 mM TCEP.

[00597] The affinity eluate was further purified by SEC. The affinity eluate was concentrated to a final volume of 550 uL by ultrafiltration at 12°C using an Amicon® (MilliporeSigma, Hayward, CA) spin concentrator with an Ultracel®-50 (MilliporeSigma, Hayward, CA) membrane. The concentrated sample was filtered using a 0.22 µm 13 mm UltraCruz® (Santa Cruz Biotechnology, Dallas, TX) PVDF syringe filter before being further purified by separation at 4°C with a flow rate of 0.4 mL/minute on a HiPrep™ 16/60 Sephacryl® S-300 (GE Healthcare, Uppsala, Sweden) column equilibrated with SEC buffer composed of 50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 0.1 mM EDTA, and 1 mM TCEP. Protein was eluted with SEC buffer and 0.75 ml fractions were collected. The earliest eluting peak, as judged by UV 280, was assumed to be high molecular weight aggregated material and the corresponding fractions were discarded. Fractions corresponding to the second peak (a shoulder on the back side of the first UV 280 peak) were pooled.

[00598] The pooled samples were exchanged into storage buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, 0.1 mM EDTA, and 1 mM TCEP by concentrating down to 200 uL with an Amicon® (MilliporeSigma, Hayward, CA) spin concentrator with an Ultracel®-50 (MilliporeSigma, Hayward, CA) membrane and then diluting 75-fold with storage buffer. The sample was concentrated a second time to 700 uL and again diluted 20-fold with storage buffer. Finally, the sample was concentrated to 4.7 mg/mL in the same ultrafiltration device and stored at -80°C.

[00599] The final purified product was analyzed spectrophotometrically to determine the final concentration of the Cascade RNP complexes and to confirm the presence of a nucleic acid component as evidenced by an absorbance at 260 nm that is greater than the absorbance at 280 nM. The concentration of the Cascade RNP complexes was determined by dividing the absorbance at 280 nm by the calculated absorbance of a 0.1% solution of the intact complex with a 1 cm path length. The predicted absorbance of a 0.1% solution of the purified complex is 2.18 cm^{-1} and was calculated by dividing the sum of the calculated extinction coefficients at 280 nm for each molecule in the complex (762240 $M^{-1}cm^{-1}$) by the sum of the molecular weights of each molecule in the complex (348952.07 g/mole).

[00600] Additionally, the final product was analyzed by SDS-PAGE with Coomassie blue staining to confirm that each Cascade protein was present in approximately the correct stoichiometry, and to assess the presence of contaminating proteins. SDS-PAGE gels were stained with Coomassie InstantBlue™ (Expedeon, San Diego, CA) stain. Gels were imaged

using a Gel doc™ EZ (Bio-Rad, Hercules, CA) imager and annotated using ImageLab (Bio-Rad, Hercules, CA) software. Each properly formed complex contained six molecules of Cas7, one molecule each of Cas6 and Cas5, and two molecules of Cse2.

[00601] C. Purification of FokI-Cas8 fusion protein

[00602] A method used to purify a fusion protein comprising a FokI nuclease fusion to the *E. coli* Type I-E Cas8 protein from bacterial over-expression pellets using immobilized metal affinity chromatography, cation exchange chromatography (CIEX), and finally size exclusion chromatography (SEC) is described herein.

[00603] The *E. coli* Type I-E FokI-Cas8 fusion protein, including a linker sequence, is described in Example 1 (SEQ ID NO:413, Table 16). The expression plasmid is described in Example 2 (SEQ ID NO:439, Table 19, FIG. 23B). Cells comprising the fusion protein were produced essentially as described in Example 4A. The Cas8 fusion protein contained a N-terminal His6 tag, a Maltose binding protein domain, a TEV cleavage site, a FokI nuclease domain, and a 30-amino acid linker. The protein was captured using immobilized metal affinity chromatography. A 50 mL conical tube containing the re-suspended cell pellets was thawed on ice. The tube was then placed in an ice water bath, and the cells were lysed by sonication using a Q500 sonicator with a ¼ inch tip (Qsonica, Newtown, CT) for a treatment cycle of three minutes with repeating cycles of 10 seconds of sonication at 40% amplitude followed by 20 seconds of rest. The lysates were clarified by centrifugation at 30,970 RCF for 30 minutes at 4°C. The clarified supernatant was then added to Hispur™ Ni-NTA (Thermo Fisher Scientific, Waltham, MA) resin, that had been pre-equilibrated with Ni-wash buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 10 mM imidazole, 5% glycerol, and 1 mM TCEP. A 2 mL bed volume of nickel affinity resin was used for 1 L of *E. coli* expression culture. After one hour of incubation at 4°C with gentle mixing, the sample was poured onto a 30 mL disposable gravity flow column (Bio-Rad, Hercules, CA), allowing the unbound material to flow through the column. The resin was washed five times with five bed volumes of Ni-wash buffer. Finally, the bound proteins were eluted with five bed volumes of Ni-elution buffer composed of 50 mM Tris pH 7.5, 100 mM NaCl, 300 mM imidazole, 5% glycerol, and 1 mM TCEP.

[00604] The nickel affinity eluate was treated with TEV protease to remove the affinity tag. TEV protease was added to the eluate at a ratio of 1:25 (w/w). The sample, including TEV, was dialyzed overnight against Ni-wash buffer using a 12 mL Slid-A-Lyzer™, 10K MWCO (Thermo Fisher Scientific, Waltham, MA) dialysis cassette.

[00605] The TEV protease and the cleaved His6-MBP fragment were removed from the dialyzed sample by Ni affinity chromatography. The dialyzed sample was poured over a clean

Hispur™ Ni-NTA (Thermo Fisher Scientific, Waltham, MA) resin column equilibrated with Ni-wash buffer. The resin was then washed with 1 column volume of Ni-NTA wash buffer. The flow through and wash were combined, concentrated, and exchanged into storage buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, and 1 mM TCEP) using an Amicon® (MilliporeSigma, Hayward, CA) spin concentrator with an Ultracel®-10 (MilliporeSigma, Hayward, CA) membrane. This sample was then frozen at -80C for storage.

[00606] The sample was thawed and further purified by cation exchange chromatography (CIEX). The sample was thawed on ice and diluted 10-fold from 0.475 mL to 4.75 mL with Cold CIEX_A buffer composed of 50 mM Tris pH 7.5, 5% glycerol, and 1 mM TCEP, resulting in a final concentration of 50 mM NaCl. A 10 mL capillary loop was used to load the sample onto a 1 mL Hitrap™ SP HP (GE Healthcare, Uppsala, Sweden) column, equilibrated with a buffer comprising CIEX_A buffer and 5% CIEX_B buffer (50 mM Tris pH 7.5, 1 M NaCl, 5% glycerol, and 1 mM TCEP). The flow rate throughout the separation was of 0.75 mL/min. The loop was emptied onto the column with 15 mL of with 5% CIEX_B buffer. The unbound sample was washed out with an additional 2 mL of 5% CIEX_B buffer. 500 µL fractions were collected as the bound proteins were eluted with an 8 mL linear gradient from 5% to 65% CIEX_B buffer. There were two major UV280 elution peaks. The four fractions corresponding to the first of those two peaks were pooled. The total pooled volume was 2 mL.

[00607] The pooled CIEX fractions were further purified by SEC. The pooled CIEX fractions were concentrated to a final volume of 0.3 mL by ultrafiltration at 12°C using an Amicon® (MilliporeSigma, Hayward, CA) spin concentrator with an Ultracel®-10 (MilliporeSigma, Hayward, CA) membrane. The concentrated sample was filtered using a 0.22 µm Ultrafree-MC GV Centrifugal (MilliporeSigma, Hayward, CA) spin filter (, and further purified by separation at 4°C with a flow rate of 0.6 mL/minute on a 10/300 Superdex™ 200 GL Increase (GE Healthcare, Uppsala, Sweden) column equilibrated with a Cas8 SEC buffer (50 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol, and 1 mM TCEP). The protein was eluted with the Cas8 SEC buffer and 0.5 ml fractions were collected. The earliest eluting peak, as judged by UV 280, was assumed to be high molecular weight aggregated material and the corresponding fractions were discarded. A second major UV 280 peak was eluted after about 14 mL. The fractions corresponding to this second peak were pooled. The pooled samples were concentrated to 40 µL with an Amicon® (MilliporeSigma, Hayward, CA) spin concentrator with an Ultracel®-3 (MilliporeSigma, Hayward, CA) membrane. The concentrated sample was stored at -80°C.

[00608] The final purified product was analyzed spectrophotometrically to determine the final concentration of the fusion protein and to confirm the absence of a significant nucleic acid component as evidenced by an absorbance at 280 nm that is greater than the absorbance at 260 nm. The concentration of the FokI-Cas8 fusion was determined by dividing the absorbance at 280 nm by the calculated absorbance of a 0.1% solution of the intact complex. The predicted absorbance of a 0.1% solution of the purified complex is 1.05 cm^{-1} and was calculated by dividing extinction coefficient at 280 nm for the FokI-Cas8 fusion (86290 $M^{-1}cm^{-1}$) by its molecular weight (82171.32 g/mole). Additionally, the final product was analyzed by SDS-PAGE gels stained with InstantBlue™ (Expedeon, San Diego, CA) stain. Gels were imaged using a Gel doc™ EZ (Bio-Rad, Hercules, CA) imager and annotated using ImageLab (Bio-Rad, Hercules, CA) software. This analysis demonstrates that the purified fusion protein was the expected size and that only a low level of contaminating proteins was present.

[00609]

Example 6

Production of dsDNA Target Sequences for Use in Biochemical Cleavage Assays

[00610] dsDNA target sequences for use in *in vitro* DNA binding or cleavage assays with Cascade or Cascade-fusion effector complexes can be produced using several different methods. This Example describes three methods to produce target sequences, including annealing of synthetic ssDNA oligonucleotides, PCR amplification of selected nucleic acid target sequences from gDNA, and/or cloning of nucleic acid target sequences into bacterial plasmids. The dsDNA target sequences were used in Cascade binding or cleavage assays.

[00611] A. Production of dsDNA target sequences by annealing synthetic ssDNA oligonucleotides

[00612] DNA oligonucleotides encoding the target region of interest comprising the target sequence that is recognized by the guide portion of CRISPR RNA, the neighboring protospacer adjacent motif (PAM), and additional 5' and 3' flanking sequences were purchased from a commercial manufacturer (Integrated DNA Technologies, Coralville, IA). Two oligonucleotides were ordered per construct, one comprising the sense strand and one comprising the nonsense strand. Table 26 lists oligonucleotide sequences that were ordered to contain a target sequence denoted J3, which is derived from bacteriophage lambda gDNA. The target and PAM sequences are flanked by 20 bp of additional sequence on both the 5' and 3' ends.

[00613]

Table 26 ssDNA Oligonucleotides		
Seq ID NO:	**Description**	**Sequence**
SEQ ID NO:466	Forward oligo, J3 target sequence	ATCATCCTCCTGACAATTTTGACAGCCCACATGGC ATTCCACTTATCACTGGCATCTTTAAAAGCCAGGA CGGTC
SEQ ID NO:467	Reverse oligo, J3 target sequence	GACCGTCCTGGCTTTTAAAGATGCCAGTGATAAGT GGAATGCCATGTGGGCTGTCAAAATTGTCAGGAG GATGAT

[00614] The oligonucleotides were annealed by mixing both oligonucleotides at equimolar concentration (10 μM) in 1X annealing buffer (6 mM HEPES, pH 7.0, and 60 mM KCl), heating at 95 °C for 2 minutes, and then slow cooling. Annealed oligonucleotides were then used directly in DNA binding and/or DNA cleavage assays with Cascade and/or Cascade-effector domain fusion RNPs.

[00615] 5' Cy5 fluorescently-labeled DNA oligonucleotides encoding the target region of interest comprising both the target sequence recognized by the guide portion of CRISPR RNA, as well as the flanking neighboring protospacer adjacent motif (PAM), and additional 5' and 3' flanking sequences, were purchased from a commercial manufacturer (Integrated DNA Technologies, Coralville, IA). Four oligonucleotides were ordered per construct, one comprising the 5' fluorescent-labeled sense strand, one comprising the 5' unlabeled sense strand, one comprising the 5' fluorescent-labeled nonsense strand, and one comprising the 5'unlabeled nonsense strand. The target and PAM sequences are flanked by 20 bp of additional sequence on both the 5' and 3' ends.

[00616] Table 27 lists oligonucleotide sequences that were ordered to contain a target sequence denoted J3, which was derived from bacteriophage lambda gDNA and a control target sequence denoted CCR5, which was derived from the human CCR5 locus.

[00617]

Table 27 ssDNA Oligonucleotides for Fluorescently Labeled dsDNA Target Sequence Formation		
SEQ ID NO:	**Description**	**Sequence**
SEQ ID NO:468	target strand J3	5'CGCCGAGCTCGAATTCTTTTGACAGCCCACATG GCATTCCACTTATCACTGGCATGGATCCTGGCTG TGGTGATG

Table 27 ssDNA Oligonucleotides for Fluorescently Labeled dsDNA Target Sequence Formation		
SEQ ID NO:	**Description**	**Sequence**
SEQ ID NO:469	non target strand J3	5'CATCACCACAGCCAGGATCCATGCCAGTGATA AGTGGAATGCCATGTGGGCTGTCAAAAGAATTC GAGCTCGGCG
SEQ ID NO:470	target strand CCR5 Site	5'CGCCGAGCTCGAATTCTTTTTAGGTACCTGGCT GTCGTCCATGCTGTGTTTGCATGGATCCTGGCTG TGGTGATG
SEQ ID NO:471	non target strand CCR5	5'CATCACCACAGCCAGGATCCATGCAAACACAG CATGGACGACAGCCAGGTACCTAAAAAGAATTC GAGCTCGGCG
SEQ ID NO:472	target strand J3	5'Cy5-CGCCGAGCTCGAATTCTTTTGACAGCCCACATGG CATTCCACTTATCACTGGCATGGATCCTGGCTGT GGTGATG
SEQ ID NO:473	non target strand J3	5'Cy5-CATCACCACAGCCAGGATCCATGCCAGTGATAA GTGGAATGCCATGTGGGCTGTCAAAAGAATTCG AGCTCGGCG
SEQ ID NO:474	target strand CCR5 Site	5'Cy5-CGCCGAGCTCGAATTCTTTTTAGGTACCTGGCTG TCGTCCATGCTGTGTTTGCATGGATCCTGGCTGT GGTGATG

[00618] The oligonucleotides were annealed by mixing a labeled and unlabeled or two labeled or two unlabeled oligonucleotides at equimolar concentration (1 μM) in 1X annealing buffer (6 mM HEPES, pH 7.0, 60 mM KCl), heating at 95 °C for 2 minutes, and then slow cooling. Annealed oligonucleotides were then used directly in DNA binding assays with Cascade and/or Cascade-effector domain fusion RNPs. Cy5 fluorescently-labeled DNA oligonucleotides were imaged with an AZURE c600 (Azure BioSystems, Dublin, CA) bioimager.

[00619] This method can be applied to produce additional labeled or unlabeled target or dual-target sequences, whereby a dual target is defined as a target that contains two protospacer sequences targeted by individual Cascade molecules, separated by an interspacer sequence.

[00620] B. Production of dsDNA target sequences by PCR amplification from gDNA

[00621] dsDNA target sequences for dual targets derived from human gDNA were produced using PCR amplification directly from gDNA template material. Specifically, PCR

reactions contained human gDNA purified from K562 cells and Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA), as well as the primers listed in Table 28, where the underlined portions correspond to primer binding sites within gDNA.

[00622]

Table 28
Primers for PCR Amplification

Seq ID NO:	Description	Sequence
SEQ ID NO:475	Forward primer to amplify Hsa07 dual-target from human gDNA	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTCCTCCCTAACCTCCACCT
SEQ ID NO:476	Reverse primer to amplify Hsa07 dual-target from human gDNA	GGAGTTCAGACGTGTGCTCTTCCGATCTTAAAGAGCCCAACCAGATGC

[00623] PCR was performed according to the manufacturer's instructions (New England Biolabs, Ipswich, MA), and the desired product DNA, 288 bp in length, was purified using a Nucleospin Gel and PCR Cleanup kit (Macherey-Nagel, Bethlehem, PA) This dsDNA was then used directly in DNA binding and/or DNA cleavage assays with Cascade and/or Cascade-effector domain fusion RNPs.

[00624] C. Production of dsDNA target sequences by cloning target sequences into bacterial plasmids

[00625] DNA oligonucleotides encoding the target region of interest comprising the target sequence, also known as the protospacer, that is recognized by the guide portion of CRISPR RNA, the neighboring protospacer adjacent motif (PAM), and additional 5' and 3' flanking sequences were purchased from a commercial manufacturer (Integrated DNA Technologies, Coralville, IA). The oligonucleotides were designed such that, when annealed, the termini regenerate sticky ends upon cleavage of their respective recognition sites by the restriction enzymes EcoRI and BlpI, or by BamHI and EcoRI. Oligonucleotides were designed to contain a single target sequence derived from the bacteriophage lambda genome, denoted J3. In addition, oligonucleotides were designed to contain two tandem target sequences derived from the bacteriophage lambda genome, denoted J3 and L3, separated from each other by a 15 bp interspacer sequence. Sequences of these oligonucleotides are listed in Table 29.

[00626]

Table 29 Oligonucleotides Comprising Target Sequences			
SEQ ID NO:	**Description**	**Restriction enzyme recognition sites**	**Sequence**
SEQ ID NO:477	Forward oligonucleotide, J3 target sequence for cloning into pACYC-Duet1	BamHI and EcoRI	GATCCATGCCAGTGATAAGTG GAATGCCATGTGGGCTGTCAA AAG
SEQ ID NO:478	Reverse oligonucleotide, J3 target sequence for cloning into pACYC-Duet1	BamHI and EcoRI	AATTCTTTTGACAGCCCACATG GCATTCCACTTATCACTGGCAT G
SEQ ID NO:479	Forward oligonucleotide, J3-15bp-L3 target sequences for cloning into pACYC-Duet1	EcoRI and BlpI	AATTCTTTTGACAGCCCACATG GCATTCCACTTATCACTGGCAT CCTAGGCCTCTCGAGATGAGTG GCAGATATAGCCTGGTGGTTCA GGCGGCGCATGC
SEQ ID NO:480	Reverse oligonucleotide, J3-15bp-L3 target sequences for cloning into pACYC-Duet1	EcoRI and BlpI	TCAGCATGCGCCGCCTGAACCA CCAGGCTATATCTGCCACTCAT CTCGAGAGGCCTAGGATGCCA GTGATAAGTGGAATGCCATGT GGGCTGTCAAAAG

[00627] The oligonucleotides contain 5'-phosphorylated ends, which were introduced by the commercial manufacturer or phosphorylated in-house using T4 polynucleotide kinase (New England Biolabs, Ipswich, MA). The oligonucleotides were then annealed at a final concentration of 1μM by mixing together equimolar amounts in annealing buffer (6 mM HEPES, pH 7.0, 60 mM KCl), heating to 95°C for 2 minutes, and then slow-cooling on the benchtop.

[00628] Separately, a pACYC-Duet1 (MilliporeSigma, Hayward, CA) plasmid was double-digested with the corresponding pair of restriction enzymes, either BamHI and EcoRI, or EcoRI and BlpI, whose sticky ends match the sticky ends formed by the termini of the hybridized oligonucleotides. The double-digested vector was separated from the removed insert using agarose gel electrophoresis.

[00629] In order to clone the hybridized oligonucleotides into the double-digested vector, the hybridized oligonucleotides were diluted to a 50 nM stock concentration, and then a 10 μL

ligation reaction was formed using hybridized oligonucleotides, the double-digested vector, and Quick Ligase (New England Biolabs, Ipswich, MA). The ligation reaction was then used to transform chemically competent *E. coli* strains, and after overnight growth on agarose plates, individual clones were isolated and grown in liquid culture to generate sufficient bacterial cultures from which to isolate plasmids. Sanger sequencing was then used to validate the desired plasmid sequence. Table 30 provides complete vector sequences for plasmids containing the J3 target sequence (SEQ ID NO:481) and plasmids containing the J3 and L3 targets sequences separated by the 15 bp interspacer sequence (SEQ ID NO:482).

[00630]

Table 30
Complete Plasmid Sequences

SEQ ID NO:	**Description of plasmid**
SEQ ID NO:481	J3 target sequence in pACYC-Duet1
SEQ ID NO:482	J3-15bp-L3 target sequences in pACYC-Duet1
SEQ ID NO:483	J3-30bp-L3 target sequences in pACYC-Duet1
SEQ ID NO:484	multi-target plasmid

[00631] Further cloning manipulations were used to generate additional double-target plasmid constructs. The 15 bp interspacer sequence of SEQ ID NO:482 contains unique AvrII and XhoI restriction sites. Thus, introduction of additional hybridized oligonucleotides into these restriction sites expands the interspacer to longer lengths, for biochemical testing with purified Cascade and Cascade-nuclease fusion RNPs. Because the crRNA-guided FokI-Cascade fusion complex targets two adjacent DNA site, dimerization of the FokI domains from adjacent DNA-bound complexes leads to DNA cleavage within the interspacer separating the two target sites. Variable interspacer lengths were designed and tested to evaluate a given interspacer length with a given tethering geometry between the FokI nuclease domain and its fused Cascade subunit protein. The complete vector sequence for a target DNA substrate containing an expanded interspacer sequence of 30 bp is given in Table 30 as SEQ ID NO:483.

[00632] In addition, the following cloning strategy provided a plasmid substrate that contains several target sequences serially connected along one large insert. A gene block was ordered from a commercial manufacturer (Integrated DNA Technologies, Coralville, IA) that contained 17 consecutive dual targets. The gene block contained 4 bp separating each dual target from a neighboring dual target, and contained 16 dual targets derived from *H. sapiens* gDNA, as well as one control dual target containing J3/L3 targets derived from the bacteriophage lambda genome. The genomic coordinates of the 16 consecutive human dual targets are shown in Table 31. The gene block was ordered with flanking SacI and SbfI

restriction sites on the ends, such that it could be cloned into SacI and SbfI sites in the pACYC-Duet1 vector. The full vector sequence of the multi-target plasmid substrate generated by cloning the gene block into pACYC-Duet1 is presented as SEQ ID NO:484 in Table 30. This multi-target sequence plasmid allowed for biochemical testing of multiple different FokI-Cascade preparations harboring crRNAs targeting one of the serially connected target sites within the plasmid.

[00633]

Table 31
Human Dual Targets

SEQ ID NO:	Target name	Gene	5' spacer target genomic coordinates	3' spacer target genomic coordinates
SEQ ID NO:485	Hsa01	PDCD1	chr2:241850348 -241850382	chr2:241850408 -241850442
SEQ ID NO:486	Hsa02	CTLA4	chr2:203870664 -203870698	chr2:203870724 -203870758
SEQ ID NO:487	Hsa03	TRAC	chr14:22509340 -22509374	chr14:22509405 -22509439
SEQ ID NO:488	Hsa04	TRAC	chr14:22509785 -22509819	chr14:22509850 -22509884
SEQ ID NO:489	Hsa05	TRAC	chr14:22513932 -22513966	chr14:22513997 -22514031
SEQ ID NO:490	Hsa06	TRAC	chr14:22515993 -22516027	chr14:22516058 -22516092
SEQ ID NO:491	Hsa07	TRAC	chr14:22516265 -22516299	chr14:22516330 -22516364
SEQ ID NO:492	Hsa08	CD52	chr1:26320402-26320436	chr1:26320467-26320501
SEQ ID NO:493	Hsa09	CTLA4	chr2:203873012 -203873046	chr2:203873077 -203873111
SEQ ID NO:494	Hsa10	CTLA4	chr2:203873195 -203873229	chr2:203873260 -203873294
SEQ ID NO:495	Hsa11	TRAC	chr14:22551630 -22551664	chr14:22551700 -22551734
SEQ ID NO:496	Hsa12	CTLA4	chr2:203872758 -203872792	chr2:203872828 -203872862
SEQ ID NO:497	Hsa13	TRAC	chr14:22551862 -22551896	chr14:22551937 -22551971
SEQ ID NO:498	Hsa14	TRBC2	chr7:142801112 -142801146	chr7:142801187 -142801221
SEQ ID NO:499	Hsa15	TRAC	chr14:22551630 -22551664	chr14:22551710 -22551744
SEQ ID NO:500	Hsa16	CTLA4	chr2:203867814 -203867848	chr2:203867894 -203867928

[00634]

Example 7

Use of Purified Cascade Complexes in Biochemical Cleavage Assays

[00635] This Example illustrates the use of FokI-Cascade fusion protein complexes in biochemical dsDNA cleavage assays. Protein reagents were compared in terms of their activity in dsDNA cleavage.

[00636] FokI–Cascade RNPs derived from the *E. coli* Type I-E Cascade system were designed, recombinantly expressed in *E. coli*, and purified for use, as outlined in Examples 1, 2, and 5. These RNPs were designed to contain either CRISPR RNAs that target the J3 and L3 target sequences derived from bacteriophage lambda gDNA, or that target an intron in the *TRAC* gene within human gDNA. Each RNP preparation is a heterogeneous mixture comprising two FokI-Cascade complexes that are otherwise identical except for the guide portion of the crRNA.

[00637] FokI-Cas8 was purified separately from the Cas8-less Cascade complex, programmed with guide polynucleotides targeted to the J3 and L3 lambda target sequences, and used in biochemical cleavage assays with J3/L3 plasmid substrates harboring the target sites in a PAM-in configuration.

[00638] A FokI-Cascade complex was reconstituted by mixing together a CasBCDE complex (produced using SEQ ID NO:440 and SEQ ID NO:446, as described in Example 2) with purified FokI-Cas8 comprising a 16-aa linker (the general FokI-Cas8 expression vector sequence is described in Example 2, SEQ ID NO:439 in Table 19; the particular 16-aa linker is in Example 1, SEQ ID NO:431 in Table 17). Reconstitution was performed in 1X Cascade Cleavage Buffer (20 mM Tris-Cl, pH 7.5, 200 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP, 5% glycerol) with CasBCDE and FokI-Cas8 both at 1 μM final concentrations.

[00639] In order to perform DNA cleavage assays, reaction mixtures were as follows. A plasmid substrate comprising the J3/L3 double-target sequence with a 30 bp interspacer (SEQ ID NO:483 in Table 30) was incubated with varying concentrations of FokI-Cascade complex (3-100 nM) in a 15 μL reaction in 1X Cascade Cleavage Buffer, with the plasmid DNA at a final concentration of 13.3 ng/μL. Reactions were incubated for 30 minutes at 37°C, after which 3 μL of 6X SDS loading dye was added. The loading dye was added to denature bound FokI-Cascade complexes. The reaction mixture components were resolved by 0.8% agarose gel electrophoresis. Gels were stained after electrophoresis with SYBR™ Safe DNA Gel Stain (Thermo Scientific, Wilmington, DE).

[00640] As a positive control, *S. pyogenes* Cas9 protein was programmed with a single-guide RNA (sgRNA) targeting a 20 bp portion of the Cascade J3 target sequence (sgRNA-J3; the spacer sequence is presented as SEQ ID NO:501). Cas9/sgRNA-J3 complexes were reconstituted by mixing Cas9 together with a 2-fold molar excess of sgRNA in 1X CCE buffer (20 mM HEPES pH 7.4, 10 mM MgCl2, 150 mM KCl, 5% glycerol). Cleavage by this Cas9/sgRNA-J3 complex was evaluated across the same concentration range (3–100 nM) by incubating reactions for 30 minutes at 37°C. Also included in the experiment were control lanes containing uncut plasmid DNA, as well as plasmid DNA linearized with the NheI restriction enzyme (New England Biolabs, Ipswich, MA). Target DNA cleavage is evidenced by a mobility shift in the plasmid, because uncut plasmid DNA is supercoiled and has a faster mobility than cleaved, linearized plasmid DNA. Nicked, open-circular plasmid DNA has a slower mobility than both supercoiled and linearized plasmid DNA.

[00641] The data obtained from these experiments demonstrate that, over the concentration range, the FokI-Cascade complex exhibited similar target DNA cleavage activity as Cas9-sgRNA. At the highest concentration tested (100 nM), the plasmid target was quantitatively linearized by the FokI-Cascade complex and Cas9-sgRNA.

[00642] FokI-Cascade complex reagents were also tested for their kinetics of target DNA cleavage. A plasmid substrate containing the J3/L3 double-target sequence with a 30 bp interspacer (SEQ ID NO:483) was incubated with 200 nM FokI–Cascade complex or 200 nM Cas9–sgRNA in a 15 µL reaction, with the plasmid DNA at a final concentration of 13.3 ng/µL. Reactions were quenched at either 0, 7, 10, 15, 20, 25, or 30 minutes, and reaction components were resolved by agarose gel electrophoresis as described above. The FokI–Cascade complex exhibited similar but slightly slower rates of target DNA cleavage activity as Cas9/sgRNA-J3 complex, with the target plasmid quantitatively linearized by the 25 minute time-point for the FokI-Cascade complex and by the 20 minute time point for the Cas9/sgRNA-J3 complex.

[00643] FokI-Cascade complex reagents were also tested for their non-specific DNA cleavage and/or nicking activity on the pACYC-Duet1 non-target plasmid substrate, versus specific DNA cleavage of a the J3/L3 double-target plasmid substrate. Table 32 contains the sequence of the pACYC-Duet1 non-target plasmid substrate used for this control (SEQ ID NO:502). Specifically, the dependence of non-specific and specific DNA target cleavage was investigated as a function of the monovalent salt concentration in the reaction buffer. Variants of the 1X Cascade Cleavage Buffer (20 mM Tris-Cl, pH 7.5, 200 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP, and 5% glycerol) were prepared, in which the NaCl concentration was dropped from 200 mM to either 150 mM, 100 mM or 50 mM, and the same cleavage reactions as

described above were performed by incubating 200 nM FokI-Cascade complex with either 13.3 ng/μL of the J3/L3 target plasmid or 13.3 ng/μL of the pACYC-Duet1 non-target plasmid. Additional control reactions were performed, in which the NaCl concentration was maintained at 100 mM, but the 5 mM MgCl2 was replaced with 10 mM EDTA, which was expected to abrogate cleavage because of the requirement of FokI for divalent metal ions for DNA cleavage. Accordingly, non-target plasmid and J3/L3 target plasmid were subjected to the following reaction conditions: –FokI-Cascade complex; +FokI-Cascade complex, 100 mM NaCl buffer + 10 mM EDTA; +FokI-Cascade complex, 50 mM NaCl buffer; +FokI-Cascade complex, 100 mM NaCl buffer; +FokI-Cascade complex, 150 mM NaCl buffer; +FokI-Cascade complex, 200 mM NaCl buffer. The data demonstrate that FokI-Cascade complex showed non-specific nicking of both the non-target and J3/L3 target plasmid at low salt concentrations <200 mM NaCl, but that at a monovalent salt concentration of 200 mM NaCl, the non-target plasmid remained intact, but the J3/L3 target plasmid was quantitatively linearized. Furthermore, buffer containing EDTA led to a complete abrogation of target cleavage, as expected.

[00644] In order to confirm that the FokI–Cascade complex cleaves the target plasmid at the expected position, that is, within the middle of the interspacer sequence separating the J3 and L3 targets, an experiment was performed in which the target plasmid was first incubated with FokI-Cascade complex, followed by incubation with the AfeI restriction enzyme (New England Biolabs, Ipswich, MA), which cleaves elsewhere in the plasmid substrate. Thus, cleavage by both FokI-Cascade 1 complex and AfeI converts the supercoiled, circular plasmid into two linear fragments migrating as distinct species on an agarose gel. Specifically, cleavage was expected to generate fragments that are 2427 bp and 1357 bp in length.

[00645] 13.3 ng/μL J3/L3 target plasmid was incubated with 200 nM FokI-Cascade 1 complex for 30 minutes, after which 1 μL of AfeI (10 Units/μL; New England Biolabs, Ipswich, MA) was added to the reaction, followed by an additional 30-minute incubation at 37°C. Reaction products were resolved by agarose gel electrophoresis, as described above. Additionally, for control experiments, the target plasmid was incubated with only FokI-Cascade 1 complex or only AfeI, and the same reactions were performed with a non-target plasmid that can be cleaved by AfeI but not by FokI-Cascade 1 complex (because the plasmid lacks the J3/L3 dual target). Table 32 contains the sequence of the pACYC-Duet1 non-target plasmid substrate used for this control (SEQ ID NO:502). Accordingly, non-target plasmid and J3/L3 target plasmid were subjected to the following reaction conditions: –AfeI/–FokI-Cascade complex; –AfeI/+FokI-Cascade complex; +AfeI/+FokI-Cascade complex; and +AfeI/–FokI-Cascade complex. The data demonstrate that FokI-Cascade complex cleaved the target plasmid

in the expected location, because co-incubation with FokI-Cascade 1 complex and AfeI lead to two linear products of the expected lengths.

[00646] In order to further confirm the sequence specificity of DNA cleavage by the FokI-Cascade complex, additional control plasmid substrates were generated that contain as follows: mutations in the PAM flanking the J3 target, mutations in the PAM flanking the L3 target, mutations in both PAMs flanking J3/L3 targets; mutations in the spacer sequence within the J3 target, mutations in the spacer sequence within the L3 target, mutations in both spacer sequences within J3/L3 targets; and the J3 target but not the L3 target, the L3 target but not the J3 target, and neither J3 nor L3 target. Accordingly, the plasmid substrates were as follows: J3 PAM mutant, L3 PAM mutant, J3/L3 PAM mutant, J3 spacer mutant, L3 spacer mutant, J3/L3 spacer mutant, non-target plasmid, J3-only target, L3-only target, and J3/L3 target plasmid. Each target was subjected to the following reaction conditions: –NdeI/–FokI-Cascade complex; +NdeI/–FokI-Cascade complex; and –NdeI/+FokI-Cascade 1 complex. Table 32 contains the sequences of all the mutated plasmid substrates described above (SEQ ID NO:502 through SEQ ID NO:510).

[00647]

Table 32
Mutated Plasmid Substrate Sequences

SEQ ID NO:	Description of plasmid
SEQ ID NO:502	pACYC-Duet1 non-target plasmid
SEQ ID NO:503	J3-30bp-L3 target plasmid, J3 PAM mutant
SEQ ID NO:504	J3-30bp-L3 target plasmid, L3 PAM mutant
SEQ ID NO:505	J3-30bp-L3 target plasmid, J3/L3 PAM mutants
SEQ ID NO:506	J3-30bp-L3 target plasmid, J3 spacer mutant
SEQ ID NO:507	J3-30bp-L3 target plasmid, L3 spacer mutant
SEQ ID NO:508	J3-30bp-L3 target plasmid, J3/L3 spacer mutants
SEQ ID NO:509	J3-only target plasmid
SEQ ID NO:510	L3-only target plasmid

[00648] DNA cleavage reactions were performed as described above, using 200 nM FokI-Cascade complex and 13.3 ng/μL plasmid substrates; control reactions to linearize each plasmid substrate were performed with NdeI (New England Biolabs, Ipswich, MA). Agarose gel electrophoresis was performed as described above. The data demonstrate that efficient double-strand beak introduction and linearization of the target plasmid is only observed for the J3/L3 target plasmids, but not for control plasmids harboring PAM or seed mutations, or only one of the two target sites.

[00649] Components for various FokI-Cascade complexes were cloned and overexpressed. RNPs produced by these components were purified and tested for biochemical DNA cleavage, in order to compare activity for different FokI-Cascade complexes. Specifically, DNA cleavage activities were compared for reconstituted FokI-Cascade complexes comprising the following: separately purified CasBCDE complex (produced using SEQ ID NO:440 and SEQ ID NO:446) and FokI-Cas8 (produced using SEQ ID NO:439); FokI-Cascade harboring the J3/L3 guide crRNAs (produced using SEQ ID NO:442 and SEQ ID NO:446); FokI-Cascade harboring an additional nuclear localization signal on either the Cas7 subunit (produced using SEQ ID NO:443 and SEQ ID NO:446) or the Cas6 subunit; FokI-Cascade harboring an additional nuclear localization signal and HA tag on either the Cas7 subunit or the Cas6 subunit; FokI-Cascade that underwent a more stringent purification involving both size exclusion chromatography (SEC) and ion exchange chromatography (IEX); and FokI-Cascade that was purified only by immobilized metal affinity chromatography (IMAC), without further clean-up.

[00650] Accordingly, non-target plasmid and J3/L3 target plasmid were subjected to the following reaction conditions: negative control; AfeI; CasBCDE+FokI-Cas8 complex; FokI-Cascade complex; FokI-Cascade (NLS-Cas6) complex; FokI-Cascade (Cas7-NLS) complex; FokI-Cascade (NLS-HA-Cas6) complex; FokI-Cascade (Cas7-HA-NLS) complex; FokI-Cascade complex (IEX, SEC clean-up); and FokI-Cascade complex (no clean-up). DNA cleavage reactions were performed with these RNP reagents as described above, using either the non-target plasmid or the consensus J3/L3 target plasmids, and reaction products were resolved by agarose gel electrophoresis. The data demonstrate that all of the RNP reagents, with one exception, exhibit nearly identical and quantitative plasmid DNA cleavage, with no background cleavage of the non-target plasmid. The sole exception was the FokI-Cascade purified without further clean-up, which exhibited more non-specific nicking activity, as seen for the lane in which it was incubated with the non-target plasmid.

[00651] Finally, using the NLS-tagged Cas7 variant of the FokI-Cascade complex as a starting point, 16 different paired guide crRNA were tested for biochemical DNA cleavage of a plasmid substrate for *H. sapiens* genomic sites Hsa01 through Hsa16 serially connected along one large insert (SEQ ID NO:484). Each pair of crRNAs contains two unique spacer sequences that correspond to two adjacent target sites in human gDNA, separated by an interspacer; the target sequences are described in SEQ ID NO:485 through SEQ ID NO:500. Table 33 contains sequences of both crRNAs within each pair that targets Hsa01 through Hsa16 gDNA

sequences; the spacer of the crRNA is underlined and in lower case, and the sequences 5' and 3' of the guide region correspond to repeat sequences from the CRISPR array.

[00652]

Table 33
crRNA Sequences

SEQ ID NO:	DNA target	crRNA sequence
SEQ ID NO:511	Hsa01-1	AUAAACCGcgggcaggcagagcuggaggccuuucaggccc GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:512	Hsa01-2	AUAAACCGggccugaggugcugccugggcauguguaaagg GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:513	Hsa02-1	AUAAACCGcacugucacccggaccucaguggcuuugccug GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:514	Hsa02-2	AUAAACCGucugugcggcaaccuacaugauggggaaugag GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:515	Hsa03-1	AUAAACCGaugagcuuguuuguagcaccaccauaaauucac GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:516	Hsa03-2	AUAAACCGuacguaaguaguggcaugugucagguggauuc GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:517	Hsa04-1	AUAAACCGaaggcauuuggaccggcagacacauaauugua GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:518	Hsa04-2	AUAAACCGagacuccagagccauccuugggaagagugcug GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:519	Hsa05-1	AUAAACCGacaagaggugugguuuccugaauucccacagug GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:520	Hsa05-2	AUAAACCGuaaguguuucuagccauccuugauuuugauca GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:521	Hsa06-1	AUAAACCGuggcuacugcucugucuccuggggauccugccu GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:522	Hsa06-2	AUAAACCGgcccauaccuucaaggaaaauuaaggcaaauaG AGUUCCCCGCGCCAGCGGGG
SEQ ID NO:523	Hsa07-1	AUAAACCGguugauuugccugcauugguguuacacagucu GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:524	Hsa07-2	AUAAACCGuaaguuguguucuucuuugccuaggccuucag GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:525	Hsa08-1	AUAAACCGgcacugccugucaacuucuacaaccuggugau GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:526	Hsa08-2	AUAAACCGuaggggccaagcagugcccagcuggggggucaa GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:527	Hsa09-1	AUAAACCGcuuucacugaaaguggagcugaugugacagaa GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:528	Hsa09-2	AUAAACCGaugugggucaaggaauuaaguuagggaauggc GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:529	Hsa10-1	AUAAACCGgcauaaaauuuuaacuugaaaagaucauuucgg GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:530	Hsa10-2	AUAAACCGgcuucaaaaauacucacauggcuauguuuuag GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:531	Hsa11-1	AUAAACCGaggggcaaugcagaggaaggagcgagggagca GAGUUCCCCGCGCCAGCGGGG

Table 33 crRNA Sequences		
SEQ ID NO:	**DNA target**	**crRNA sequence**
SEQ ID NO:532	Hsa11-2	AUAAACCGgaggugaaagcugcuaccaccucugugccccc GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:533	Hsa12-1	AUAAACCGgcugaaauugcuuuucacauucuggcucuguu GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:534	Hsa12-2	AUAAACCGagaguccauauuucaauuuccaagagcugagg GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:535	Hsa13-1	AUAAACCGugcacagccaggggaggcugcagcagccuugc GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:536	Hsa13-2	AUAAACCGauggaucuucaguggguucucuugggcucuag GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:537	Hsa14-1	AUAAACCGccuguggccaggcacaccagugUGGCCUU UUGGAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:538	Hsa14-2	AUAAACCGgaggugcacaguggggucagcacagacccgca GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:539	Hsa15-1	AUAAACCGaggggcaaugcagaggaaggagcgagggagca GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:540	Hsa15-2	AUAAACCGcugcuaccaccucugugccccccggcaaugcG AGUUCCCCGCGCCAGCGGGG
SEQ ID NO:541	Hsa16-1	AUAAACCGgacuuuauauagauagcuuugaucccagauau GAGUUCCCCGCGCCAGCGGGG
SEQ ID NO:542	Hsa16-2	AUAAACCGguuuugcucuacuuccugaagaccugaacacc GAGUUCCCCGCGCCAGCGGGG

[00653] After the 16 FokI–Cascade complexes were purified, cleavage reactions were performed as described above, wherein the FokI-Cascade complexes were incubated with the plasmid substrate containing *H. sapiens* genomic sites Hsa01 through Hsa16, and the reaction products were resolved by agarose gel electrophoresis. The data demonstrate that, of the 16 RNP reagents, 14/16 (Hsa03–Hsa16) exhibited nearly quantitative DNA cleavage, as evidenced by conversion of the supercoiled, circular plasmid substrate into the cleaved, linear form. Only constructs Hsa01 and Hsa02 showed partial nicking activity. Furthermore, the data demonstrated that the FokI-Cascade complexes were effectively programmed using the designed 16 paired gRNAs to target therapeutically relevant *H. sapiens* genes.

[00654]

Example 8

Introduction of FokI-Cascade RNP Complexes into Target Cells

[00655] This Example illustrates the design and delivery of *E. coli* Type I-E Cascade complexes comprising FokI fusion proteins to facilitate genome editing in human cells and describes their delivery into target cells as pre-assembled Cascade RNP complexes.

[00656] A. Production of Cascade RNP complexes comprising FokI for transformation into cells

[00657] Minimal CRISPR arrays were designed to target eight distinct loci in the human genome. Each minimal CRISPR array contained two spacer sequences, both of which were flanked by CRISPR repeat sequences. The two spacer sequences targeted loci in the genome separated by 30 bp (i.e., a 30 bp interspacer region), and each spacer was designed to bind a target sequence adjacent to an AAG or ATG protospacer adjacent motif (PAM) sequence in the target cell genome. Plasmid vectors containing each minimal CRISPR array were produced by ligating annealed oligonucleotides (Integrated DNA Technologies, Coralville, IA) into a pACYC-Duet1 (MilliporeSigma, Hayward, CA) vector backbone for bacterial expression.

[00658] Overlapping primers to produce selected spacers in minimal CRISPR arrays are set forth in Table 34, and the sequences of the primers are described in Table 35.

[00659]

Table 34
Overlapping Primers for Generation of Minimal CRISPR arrays

Component	Gene target	Primers
Hsa03 Minimal CRISPR array	TRAC intron	A,B
Hsa04 Minimal CRISPR array	TRAC intron	C,D
Hsa05 Minimal CRISPR array	TRAC intron	E,F
Hsa06 Minimal CRISPR array	TRAC intron	G,H
Hsa07 Minimal CRISPR array	TRAC intron	I,J
Hsa08 Minimal CRISPR array	CD52 exon	K,L
Hsa09 Minimal CRISPR array	CTLA4 exon	M,N
Hsa10 Minimal CRISPR array	CTLA4 exon	O,P

[00660]

Table 35
DNA Primer Sequences

SEQ ID NO:	Oligo-nucleotide	Sequence
SEQ ID NO:543	A	/5Phos/ACCGATGAGCTTGTTTGTAGCACCACCATAATTC ACGAGTTCCCCGCGCCAGCGGGGATAAACCGTACGTA AGTAGTGGCATGTGTCAGGTGGATTC
SEQ ID NO:544	B	/5Phos/ACTCGAATCCACCTGACACATGCCACTACTTACG TACGGTTTATCCCCGCTGGCGCGGGGAACTCGTGAATT ATGGTGGTGCTACAAACAAGCTCAT
SEQ ID NO:545	C	/5Phos/ACCGAAGGCATTTGGACCGGCAGACACATAATT GTAGAGTTCCCCGCGCCAGCGGGGATAAACCGAGACT CCAGAGCCATCCTTGGGAAGAGTGCTG
SEQ ID NO:546	D	/5Phos/ACTCCAGCACTCTTCCCAAGGATGGCTCTGGAGT CTCGGTTTATCCCCGCTGGCGCGGGGAACTCTACAATT ATGTGTCTGCCGGTCCAAATGCCTT

Table 35 DNA Primer Sequences		
SEQ ID NO:	**Oligo-nucleotide**	**Sequence**
SEQ ID NO:547	E	/5Phos/ACCGACAAGAGGTGTGTTTCCTGAATTCCCACA GTGGAGTTCCCCGCGCCAGCGGGGATAAACCGTAAGT GTTTCTAGCCATCCTTGATTTTGATCA
SEQ ID NO:548	F	/5Phos/ACTCTGATCAAAATCAAGGATGGCTAGAAACAC TTACGGTTTATCCCCGCTGGCGCGGGGAACTCCACTGT GGGAATTCAGGAAACACACCTCTTGT
SEQ ID NO:549	G	/5Phos/ACCGTGGCTACTGCTCTGTCTCCTGGGATCCTGC CTGAGTTCCCCGCGCCAGCGGGGATAAACCGGCCCAT ACCTTCAAGGAAAATTAAGGCAAATA
SEQ ID NO:550	H	/5Phos/ACTCTATTTGCCTTAATTTTCCTTGAAGGTATGG GCCGGTTTATCCCCGCTGGCGCGGGGAACTCAGGCAG GATCCCAGGAGACAGAGCAGTAGCCA
SEQ ID NO:551	I	/5Phos/ACCGGTTGATTTGCCTGCATTGGTGTTACACAGT CTGAGTTCCCCGCGCCAGCGGGGATAAACCGTAAGTTG TGTTCTTCTTTGCCTAGGCCTTCAG
SEQ ID NO:552	J	/5Phos/ACTCCTGAAGGCCTAGGCAAAGAAGAACACAAC TTACGGTTTATCCCCGCTGGCGCGGGGAACTCAGACTG TGTAACACCAATGCAGGCAAATCAAC
SEQ ID NO:553	K	/5Phos/ACCGGCACTGCCTGTCAACTTCTACAACCTGGTG ATGAGTTCCCCGCGCCAGCGGGGATAAACCGTAGGGG CCAAGCAGTGCCCAGCTGGGGGTCAA
SEQ ID NO:554	L	/5Phos/ACTCTTGACCCCCAGCTGGGCACTGCTTGGCCCC TACGGTTTATCCCCGCTGGCGCGGGGAACTCATCACCA GGTTGTAGAAGTTGACAGGCAGTGC
SEQ ID NO:555	M	/5Phos/ACCGCTTTCACTGAAAGTGGAGCTGATGTGACA GAAGAGTTCCCCGCGCCAGCGGGGATAAACCGATGTG GGTCAAGGAATTAAGTTAGGGAATGGC
SEQ ID NO:556	N	/5Phos/ACTCGCCATTCCCTAACTTAATTCCTTGACCCAC ATCGGTTTATCCCCGCTGGCGCGGGGAACTCTTCTGTC ACATCAGCTCCACTTTCAGTGAAAG
SEQ ID NO:557	O	/5Phos/ACCGGCATAAAATTTAACTTGAAAAGATCATTT CGGGAGTTCCCCGCGCCAGCGGGGATAAACCGGCTTC AAAAATACTCACATGGCTATGTTTTAG
SEQ ID NO:558	P	/5Phos/ACTCCTAAAACATAGCCATGTGAGTATTTTTGAA GCCGGTTTATCCCCGCTGGCGCGGGGAACTCCCGAAAT GATCTTTTCAAGTTAAATTTTATGC
SEQ ID NO:559	Q	CACTCTTTCCCTACACGACGCTCTTCCGATCTAGCCTGG AAAGACACAAAGC
SEQ ID NO:560	R	GGAGTTCAGACGTGTGCTCTTCCGATCTCAGCCATCCT TTCCACCTAA
SEQ ID NO:561	S	CACTCTTTCCCTACACGACGCTCTTCCGATCTATGCTGC AGGCTTTATGCTT
SEQ ID NO:562	T	GGAGTTCAGACGTGTGCTCTTCCGATCTTTAGGCCTGC CTGACTTCTC
SEQ ID NO:563	U	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGGAAG AAGACCAACAAGAGG
SEQ ID NO:564	V	GGAGTTCAGACGTGTGCTCTTCCGATCTTTCAAGGGAA GAAGCCATTG

Table 35 DNA Primer Sequences		
SEQ ID NO:	**Oligo-nucleotide**	**Sequence**
SEQ ID NO:565	W	CACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGCAGGAATTGGATGAAA
SEQ ID NO:566	X	GGAGTTCAGACGTGTGCTCTTCCGATCTAACCTGAGATGACTGCCCAT
SEQ ID NO:567	Y	CACTCTTTCCCTACACGACGCTCTTCCGATCTTTCCTCCCTAACCTCCACCT
SEQ ID NO:568	Z	GGAGTTCAGACGTGTGCTCTTCCGATCTTAAAGAGCCCAACCAGATGC
SEQ ID NO:569	A2	CACTCTTTCCCTACACGACGCTCTTCCGATCTGTCTCAGCCTTAGCCCTGTG
SEQ ID NO:570	B2	GGAGTTCAGACGTGTGCTCTTCCGATCTCCCACTGCAAGTACAAGGGT
SEQ ID NO:571	C2	CACTCTTTCCCTACACGACGCTCTTCCGATCTGGATGCGGAACCCAAATTA
SEQ ID NO:572	D2	GGAGTTCAGACGTGTGCTCTTCCGATCTTAGTCTTCTCCCTCGCTCCC
SEQ ID NO:573	E2	CACTCTTTCCCTACACGACGCTCTTCCGATCTTGCAGCATTATGATGTGGGT
SEQ ID NO:574	F2	GGAGTTCAGACGTGTGCTCTTCCGATCTCAACCTTTAGCATCACTGGCT
SEQ ID NO:575	G2	CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTC
SEQ ID NO:576	H2	AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTTCCCTACACGACG

[00661] The design of bacterial expression vectors for production of Cascade RNP complexes is detailed in Example 2. In brief, each *cas* gene was expressed from a single operon, and the coding sequences for the *cas* genes were arranged in the order of *cas8–cse2–cas7–cas5–cas6*. The FokI moiety was attached by a 30-aa linker to Cas8, and a nuclear localization signal (NLS) was attached to the N-terminus of FokI-Cas8 (FokI-Cascade complex) and the N-terminus of Cas6 (hereafter referred to as FokI-Cascade-NLS-Cas6 complex, SEQ ID NO:577).

[00662] FokI-Cascade-NLS-Cas6 complexes were purified as assembled complexes from *E. coli* essentially as described in Example 5A.

[00663] B. Transfection of Cascade RNP complexes comprising FokI into eukaryotic cells

[00664] HEK293 cells (ATCC, Manassas, VA) were cultured in suspension in DMEM medium supplemented with 10% FBS and 1x Antibiotic-Antimycotic Solution (Mediatech, Inc., Manassas, VA) at 37°C, 5% CO_2 and 100% humidity. HEK293 cells were transfected

using the Nucleofector® 96-well Shuttle System (Lonza, Allendale, NJ). Prior to nucleofection, 5 µl of FokI-Cascade RNPs were transferred to individual wells of a 96-well plate. Each well contained ~225-500 pmol of FokI-Cascade-NLS-Cas6 complexes, depending on the RNP. HEK293 cells were transferred to a 50 ml conical centrifuge tube and centrifuged at 200xG for 3 minutes. The media was aspirated and the cell pellet was washed in calcium and magnesium-free PBS. The cells were centrifuged once more and re-suspended in Nucleofector SF (Lonza, Allendale, NJ) buffer at a concentration of $1x10^7$ cells/ml. 20 µl of this cell suspension was added to the FokI-Cascade-NLS-Cas6 complexes in the 96-well plate, mixed, and then the entire volume was transferred to a 96-well Nucleocuvette™ (Lonza, Allendale, NJ) plate. The plate was then loaded into the Nucleofector™ 96-well Shuttle™ System (Lonza, Allendale, NJ)and cells were nucleofected using the 96-CM-130 Nucleofector™ program (Lonza, Allendale, NJ). Immediately following nucleofection, 80 µl of complete DMEM medium was added to each well of the 96-well Nucleocuvette™ (Lonza, Allendale, NJ) plate. The entire contents of the well were then transferred to a 96-well tissue culture plate containing 100 µl of complete DMEM medium. The cells were cultured at 37°C, 5% CO_2 and 100% humidity for ~72 hours.

[00665] After ~72 hours, the HEK293 cells were centrifuged at 500xG for 5 minutes and the medium was removed. The cells were washed in calcium and magnesium-free PBS. The cell pellets were then re-suspended in 50 µl of QuickExtract DNA Extraction solutions (Epicentre, Madison, WI). The gDNA samples obtained were then incubated at 37°C for 10 minutes, 65°C for 6 minutes, and 95°C for 3 minutes to stop the reaction. gDNA samples were then diluted with 50µl of water and stored at -20°C for subsequent deep sequencing analysis.

[00666] C. Deep sequencing of gDNA from transfected cells

[00667] Using the isolated gDNA, a first PCR was performed using Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA) at 1x concentration, primers at 0.5 µM each, 3.75 µL of gDNA in a final volume of 10 µL and amplified 98°C for 1 minutes, 35 cycles of 10 seconds at 98°C, 20 seconds at 60°C, 30 seconds at 72°C, and a final extension at 72°C for 2 minutes. PCR reaction was diluted 1:100 in water. Target-specific primers are shown in Table 36. The target-specific primers contained Illumina-compatible sequences so that the amplification products could be analyzed using a MiSeq Sequencer (Illumina, San Diego, CA).

[00668]

Table 36 Target-specific Primers Used for Sequencing	
Target	**Oligonucleotide***
Hsa03 on-target	Q,R
Hsa04 on-target	S,T
Hsa05 on-target	U,V
Hsa06 on-target	W,X
Hsa07 on-target	Y,Z
Hsa08 on-target	A2,B2
Hsa09 on-target	C2,D2
Hsa10 on-target	E2,F2

*DNA primer sequences are shown in Table 35

[00669] A second "barcoding" PCR was set up such that each target was amplified with primers (G2 and H2 in Table 35) that each contained unique 8-bp indices (denoted by "NNNNNNNN" in the primer sequence (see SEQ ID NO:575 and SEQ ID NO:576), thus allowing de-multiplexing of each amplicon during sequence analysis.

[00670] The second PCR was performed using Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA) at 1x concentration, primers at 0.5 μM each, 1μL of 1:100 diluted first PCR, in a final volume of 10 μL and amplified 98°C for 1 minute, 12 cycles of 10 seconds at 98°C, 20 seconds at 60°C, 30 seconds at 72°C, and a final extension at 72°C for 2 minutes. PCR reactions were pooled into a single microfuge tube for SPRIselect bead (Beckman Coulter, Pasadena, CA)-based cleanup of amplicons for sequencing.

[00671] To pooled amplicons, 0.9x volumes of SPRIselect beads were added, mixed and incubated at room temperature for 10 minutes. The microfuge tube was placed on a magnetic tube stand (Beckman Coulter, Pasadena, CA) until solution had cleared. Supernatant was removed and discarded, and the residual beads were washed with 1 volume of 85% ethanol, and incubated at room temperature (RT) for 30 seconds. After incubation, ethanol was aspirated and beads were air dried at room temperature for 10 minutes. The microfuge tube was then removed from the magnetic stand and 0.25x volumes of water was added to the beads, mixed vigorously, and incubated for 2 min. at RT. The microfuge tube was spun in a microcentrifuge to collect the contents of the tube, and was then returned to the magnet, incubated until solution had cleared, and the supernatant containing the purified amplicons were dispensed into a clean microfuge

tube. The purified amplicon library was quantified using the Nanodrop™ 2000 (Thermo Scientific, Wilmington, DE) system.

[00672] The amplicon library was normalized to 4 nM concentration as calculated from optical absorbance at 260nm (Nanodrop™ 2000 (Thermo Scientific, Wilmington, DE) system) and size of the amplicons. Library was analyzed on MiSeq Sequencer (Illumina, San Diego, CA) with MiSeq Reagent Kit v2, 300 cycles (Illumina, San Diego, CA), with two 151-cycle paired-end run plus two eight-cycle index reads.

[00673] D. Deep sequencing data analysis

[00674] The identity of products in the sequencing data was analyzed based upon the index barcode sequences adapted onto the amplicons in the second round of PCR. A computational script executing the following tasks was used to process the MiSeq (Illumina, San Diego, CA) data:

[00675] Reads were aligned to the human genome (build GRCh38/38) using Bowtie (bowtie-bio.sourceforge.net/index.shtml) software.

[00676] Aligned reads were compared to wild-type loci; reads not aligning to any part of the loci were discarded.

[00677] Reads matching wild-type sequence were tallied. Reads with indels (surrounding 10 bp from the FokI-Cascade RNP expected cut site) were categorized by indel type and tallied.

[00678] Total indel reads were divided by the sum of wild-type reads and indel reads to give percent-mutated reads.

[00679] FIG. 28 shows genome editing (FIG. 28, vertical axis, "% editing") as a function of FokI-Cascade-NLS-Cas6 complex nucleofection (n=1) (FIG. 27, horizontal axis, Hsa3, Hsa4, Hsa5, Hsa6, Hsa7, Hsa8, Hsa9, and Hsa10). In FIG. 28, the open bars are the negative controls and the black bars are addition of FokI-Cascade-NLS-Cas6 complexes). FokI-Cascade-NLS-Cas6 complexes induced editing at all eight loci. Editing ranged from ~0.2-5% indels, and indels were centered around the predicted cut site, in the middle of the interspacer region.

[00680]

Example 9

Introduction of Plasmids Encoding Components of FokI-Cascade RNP Complexes into Target Cells

[00681] This Example illustrates the design and delivery of *E. coli* Type I-E Cascade complexes comprising FokI fusion proteins to facilitate genome editing in human cells. This Example also describes the delivery of plasmid vectors expressing Cascade complex components into eukaryotic cells.

[00682] A. Production of a vector encoding FokI-Cascade RNP components to be transfected into target cells

[00683] A minimal CRISPR array was designed to target the TRAC locus in the human genome. The minimal CRISPR array contained two spacer sequences, both of which were flanked by CRISPR repeat sequences, as described in Examples 1 and 3. The two spacer sequences targeted loci in the genome separated by 30 bp and each spacer was complementary to a genomic sequence adjacent to an AAG PAM sequence. The plasmid vector containing the minimal CRISPR array was produced by ligating annealed oligonucleotides (Integrated DNA Technologies, Coralville, IA) encoding a CRISPR repeat flanked by two spacer sequences into a mammalian expression vector with two CRISPR repeat sequences. The resulting plasmid contained a "repeat-spacer-repeat-spacer-repeat" that expressed two guides from the human U6 (hU6) promoter (SEQ ID No:454).

[00684] FokI-Cascade RNP protein component-encoding genes were cloned into plasmid vectors containing CMV promoters to enable delivery and expression in mammalian cells. *Cas* genes were cloned into separate plasmids (SEQ ID NO:448 through SEQ ID NO:451and SEQ ID NO:453) or in a single plasmid as a polycistronic construct with each gene linked via 2A viral peptide "ribosome-skipping" sequences (in SEQ ID NO:455). FokI-Cascade RNP complexes were delivered into eukaryotic cells via two different methods: *cas* genes and the minimal CRISPR array were supplied on separate plasmids (six-plasmid delivery system, SEQ ID NO:448 through SEQ ID NO:451, SEQ ID NO:453 and SEQ ID NO:454), or one plasmid encoding all *cas* genes as a polycistronic construct and a second plasmid encoding the minimal CRISPR array (two-plasmid delivery system, SEQ ID NO:454 and SEQ ID NO:455).

[00685] B. Transfection of plasmid(s)-encoding FokI-Cascade RNP complexes

[00686] Transfection conditions for the six-plasmid delivery system and two-plasmid delivery systems were performed as detailed in Example 8B with the following modifications. Prior to nucleofection, 5 µl of plasmid vector solution was transferred to individual wells of a 96-well plate. The six-plasmid delivery system was initially tested by examining the necessity of each component for genome editing. More specifically, plasmid "cocktails" were added to each well such that there was a constant amount (420 ng) of five plasmids and a variable amount of the sixth plasmid (either 0 ng, 70 ng, 700 ng, or 1,400 ng). Next, the six plasmid delivery system and the two-plasmid delivery system were compared by nucleofecting in a fixed amount (3.5 µg) of total plasmid DNA while varying the ratio of minimal CRISPR array plasmid to *cas*-encoding plasmid(s). Finally, lysate was harvested ~72 hours after nucleofection for subsequent deep sequencing analysis.

[00687] C. Deep sequencing of gDNA from transfected cells and data analysis

Deep sequencing was performed as detailed in Example 8C, but only using target-specific primers Y and Z from Table 36.

[00688] D. Deep sequencing data analysis

[00689] Deep sequencing data analysis was performed as detailed in Example 8D. FIG. 29 shows genome editing (FIG. 29, vertical axis, "% editing") at the TRAC locus as a function of each FokI-Cascade component in the six-plasmid delivery strategy (n=1) (FIG. 29, horizontal axis, Guide, FokI-Cas8, Cse2, Cas7, Cas5, Cas6, and reference sample). In FIG. 29, open bars represent 0 ng of the FokI-Cascade component, stippled bars represent 70 ng of the FokI-Cascade component, square pattern bars represent 700 ng of the FokI-Cascade component, and striped bars 1,400 ng of the FokI-Cascade component (bar order on the horizontal axis is left to right, respectively, for each FokI-Cascade component). As is shown, editing was abolished or dramatically reduced (in the case of Cse2) if a given component was lacking. This confirms that each Cascade component is necessary for editing via plasmid delivery.

[00690] FIG. 30 shows data comparing genome editing with the six-plasmid delivery system or the two-plasmid delivery system. FIG. 30 shows genome editing (FIG. 30, vertical axis, "% editing") at the target locus as a function of varying concentrations of each component of the six-plasmid (FIG. 30, open bars) and two-plasmid (FIG. 30, black bars) systems (FIG. 30, bar order on the horizontal axis is left to right, respectively, six-plasmid system and two-plasmid system). Numerical groupings along the horizontal axis refer to amounts of components: top line=total plasmid in ng, second line=minimal CRISPR array plasmid in ng, and third line=Cas-encoding plasmids in ng (e.g., first numerical grouping: top line=total plasmid, 3500 ng; second line=minimal CRISPR array plasmid, 0 ng; and third line=Cas-encoding plasmids, 3500 ng).

[00691] Across both methods, the highest levels of editing were achieved with the highest ratio of *cas*:minimal CRISPR array plasmids. Additionally, the polycistronic plasmid enabled higher levels of editing, potentially due to increased transcription per μg of plasmid.

[00692]

Example 10

Circular Permutations of Cascade Subunit Proteins

[00693] This Example illustrates in silico design, cloning, expression, and purification of a circularly-permuted (cp) *E. coli* Type I-E Cas7 protein using a structure-guided modelling approach.

[00694] A. In silico design

[00695] An *E. coli* Type I-E Cas7 protein (SEQ ID NO:18) was circularly permuted using a structure-guided approach based on the *E. coli* Cascade crystal structure 5H9E.pdb (www.rcsb.org/pdb/; Hayes, R.P, et al., Nature 530(7591):499-503 (2016)). The native Cas7 N-terminus and C-terminus were connected with a two-amino acid peptide linker having the sequence glycine-serine (G-S). The polypeptide sequence of this circularized Cas7 was opened at the position corresponding to the peptide bond between residues 301 and 302 in wild-type Cas7 polypeptide sequence to form a new N-terminus (residue 302) and a new C-terminus (residue 301), resulting in a circular permuted version of Cas7 protein (cp-Cas7 V1 protein). The new N-terminus and new C-terminus were designed to be positioned for connection with a fusion protein or linker region without disturbing the Cas7 protein fold or the Cascade complex assembly. A methionine residue was added to the new N-terminus (i.e., the amino acid residue corresponding to residue 302 of the wild-type Cas7 protein) of the cp-Cas7 V1 protein (SEQ ID NO:578).

[00696] A second cp-Cas7 protein, cp-Cas7 V2 protein, was similarly engineered using the G-S linker. The N-terminus and C-terminus of the cp-Cas7 V2 protein correspond to residues 338 and 339, respectively, in the wild-type Cas7 sequence. The new N-terminus and new C-terminus were designed to be positioned for connection with a fusion protein or linker region without disturbing the Cas7 protein fold or the Cascade complex assembly. A methionine residue was added to the N-terminus (i.e., the amino acid residue corresponding to residue 339 of the wild-type Cas7 protein) of the cp-Cas7 V2 protein (SEQ ID NO:579).

[00697] B. Cloning, expression, and purification of Cascade complexes comprising cp-Cas7

[00698] DNA coding sequences of the in silico designed polypeptide sequences of cp-Cas7 V1 protein and cp-Cas7 V2 protein were codon-optimized for expression in *E. coli*.

[00699] These DNA coding sequences were provided to a commercial manufacturer (GenScript, Piscataway, NJ) for synthesis. The DNA sequences were individually introduced into a Cascade-operon expression vector (Table 19; SEQ ID NO:441) to replace the wild-type Cas7 protein in the expression vector as described in Example 2.

[00700] Each expression vector was transfected into *E. coli* BL21 Star™ (Thermo Fisher Scientific, Waltham, MA) cells with a second vector encoding a guide RNA for the J3 target (SEQ ID NO:444) set forth in Table 20, as described in Example 2. Cells were cultured as described in Example 4B. *E. coli* Type I-E Cascade complexes containing Cas5, Cas6, cp-Cas7 V1, Cse2, and Cas8 proteins, as well as guide RNA/target J3; and Cas5, Cas6, cp-Cas7 V2,

Cse2, and Cas8 proteins as well as guide RNA/target J3, were purified as described in Example 5A.

[00701] Purification of the Cascade complexes comprising the cp-Cas7 variants demonstrate that circularly-permuted Type I-E CRISPR-Cas subunit proteins can be successfully used to form Cascade complexes having essentially the same composition (based on molecular weight) as Cascade complexes comprising wild-type proteins.

[00702] C. EMSA (electrophoretic mobility shift assays) of Cascade/cp-Cas7 and J3 target

[00703] Purified Cascade/cp-Cas7 complexes were purified as described in this Example and subjected to an EMSA to demonstrate specific binding to their respective target sequence. Briefly, Cascade/cp-Cas7 and Cascade/wt-Cas7 were purified and concentrated to 10 mg/mL. Cy5 double-stranded target DNA was produced essentially as described in Example 6A and diluted to 1 μM in TE buffer (J3 target SEQ ID NO:469 and SEQ ID NO:472 and CCR5 target SEQ ID NO:474 and SEQ ID NO:470). Cascade complexes and labeled double-stranded target DNA were incubated for 30 min at 37°C at different protein/target ratios. Immediately following the incubation, 2 μl of 50% glycerol was added to the samples and they were loaded on a 5% native PAA gel. Gels were run at 4°C at 70V for 90 min in 0.5xTBE buffer and imaged on an AZURE c600 Bioimager (Azure BioSystems, Dublin, CA) and the bands were quantitated. The data are presented in Table 37.

[00704]

Table 37
Results of Cascade/cp-Cas7 V2 EMSA

Cascade ID and guide	Target DNA	Cascade:dsDNA ratio	Gel shift %
Cascade/wt-Cas7 J3	J3	6.7	44
Cascade/cp-Cas7 V2 J3	J3	6.7	90
Cascade/wt-Cas7 J3	CCR5	6.7	LOD*
Cascade/cp-Cas7 V2 J3	CCR5	6.7	LOD

*LOD = below the limit of detection

[00705]

Example 11

Cascade Subunit Fusion Proteins

[00706] A. Cascade subunit fusion with FokI

[00707] This Example illustrates in silico design, cloning, expression, and purification of a *E. coli* Type I-E Cas8 protein fused to a FokI nuclease domain to confer nuclease activity to the Cascade complex.

[00708] *E. coli* Type I-E Cas8 was fused N-terminally with a *Flavobacterium okeanokoites* FokI nuclease domain (GenBank no. AAA24927.1). The FokI nuclease domain comprises residues contained in the *Sharkey* variant described by Guo, et al. (Guo, J., et al., J. Mol. Biol. 400:96-107 (2010)), and catalyzes double-stranded DNA cleavage upon homo-dimerization. The amino acid sequence for the FokI nuclease (SEQ ID NO:580) contained residues Q384 to F579 (GenBank no. AAA24927.1) and had the following point mutations: E486Q, L499I, and D469N. Briefly, the FokI *Sharkey* nuclease domain (SEQ ID NO:581) was fused N-terminal to Cas8 using a linker sequence (SEQ ID NO:582). For purification purposes, a hexahistine tag (His6, SEQ ID NO:583), followed by a MBP tag (SEQ ID NO:584), followed by a TEV protease cleavage sequence (SEQ ID NO:585), a nuclear localization signal (NLS, SEQ ID NO:586), and a GGS linker were appended N-terminal to residue 384 of FokI. The final construct comprised NH3-His6-MBP-TEV-NLS-GGS-FokI*Sharkey*-30aa-linker-Cas8-COOH in the protein sequence (SEQ ID NO:413).

[00709] In silico designed DNA sequences were provided to a commercial manufacturer (GenScript, Piscataway, NJ) for synthesis. The DNA sequences were cloned into a pET expression (MilliporeSigma, Hayward, CA) family vector backbone, which confers kanamycin resistance due to the presence of the *kanR* gene as described in Example 2 resulting in a vector carrying NH3-His6-MBP-TEV- NLS-GGS-FokI*Sharkey*-30aa-linker-Cas8-COOH (SEQ ID NO:439).

[00710] The *E. coli* Type I-E Cascade H3-His6-MBP-TEV- NLS-GGS-FokI*Sharkey*-30aa-linker-Cas8-COOH (SEQ ID NO:439) was expressed and purified as described in Example 4B and Example 5C. The protein sequence after TEV cleavage comprises NH3-NLS-GGS-FokI*Sharkey*-30aa-linker-Cas8-COOH (SEQ ID NO:587).

[00711] Similarly, a Fok1-Cas8 fusion protein was constructed in a vector that carries NLS-FokI-linker-Cas8_His6–HRV3C–Cse2_Cas7_Cas5_Cas6 as described in Examples 1 and 2 (SEQ ID NO:442). Each expression vector was transfected into *E. coli* BL21 Star™ (Thermo Fisher Scientific, Waltham, MA) cells with a second vector encoding a guide RNA for the J3

target (SEQ ID NO:444), as described in Example 2. This construct was expressed and purified as described in Example 4B and Example 5A. Purification of the Cascade complexes comprising the fused FokI-Cas8 variants demonstrate that nuclease fused Type I-E CRISPR-Cas subunit proteins can be successfully used to form Cascade complexes having essentially the same composition (based on molecular weight) as Cascade complexes comprising wild-type proteins. FokI-Cas8 fusions were successfully used for biochemical cleavage of target nucleic acid (Example 7) and for in-cell cleavage of genomic sequences in eukaryotic cells (Example 8D and Example 9D).

[00712] Table 38 lists further examples of Cas subunit protein-enzyme fusions. In Table 38, APOBEC corresponds to a gene that is member of the cytidine deaminase pathway (human APOBEC I Genbank no. AB009426, human APOBEC 3F Genbank no. CH471095, human APOBEC 3G Genbank no. CR456472, rat APOBEC UCSC genome browser ID RGD:2133 rat); AID corresponds to an activation-induced cytidine deaminase (Genbank no. AY536516); PmCDA1 is an AID ortholog (see, e.g., Nishida, et al., Science 16:353 (2016); Iwamatsu, et al., J. Biochem. 110:151-158 (1991)); PvuIIHIFIT46G is a PvuII high fidelity variant T46G (see, e.g., Fonfara, et al., Nucleic Acids Res. 40:847-860 (2012)); PvuIIsinglechainT46G is described in pdbID 3KSK); I-TevI is a site-specific, sequence-tolerant homing endonuclease from bacteriophage T4 and comprises an N-terminal catalytic domain as well as a C-terminal DNA-binding domain (the domains are connected by a long, flexible linker) (see, e.g., Van Roey, et al., EMBO J. 20:3631-3637 (2001)); BcnI (see, e.g., Sokolowska, et al., J. Mol. Biol. 369:722-734 (2007)); and MvaI (see, e.g., Kaus-Drobek, et al., Nucleic Acids Res. 35:2035-2046 (2007)) are restriction enzymes.

[00713]

Table 38
Other Enzyme Fusions (Such as Nucleases and Cytidine Deaminases) with Cas8

SEQ ID NO:	Enzyme	Fusion to Cas8
SEQ ID NO:593	Cas8_rAPOBEC1	C terminal
SEQ ID NO:594	Cas8_AID	C terminal
SEQ ID NO:595	Cas8_PmCDA1	C terminal
SEQ ID NO:596	Cas8_Human APOBEC1	C terminal
SEQ ID NO:597	Cas8_APOBEC3F	C terminal
SEQ ID NO:598	Cas8_APOBEC3G	C terminal

Table 38 Other Enzyme Fusions (Such as Nucleases and Cytidine Deaminases) with Cas8		
SEQ ID NO:	**Enzyme**	**Fusion to Cas8**
SEQ ID NO:599	PvuIIHIFIT46G	N terminal
SEQ ID NO:600	PvuIIsinglechainT46G	N terminal
SEQ ID NO:601	I-TevI1-169Q158R	N terminal
SEQ ID NO:602	I-TevI1-169	N terminal
SEQ ID NO:603	BcnI singlechain	N terminal
SEQ ID NO:604	MvaI singlechain	N terminal
SEQ ID NO:605	DNaseI	N terminal, C terminal
SEQ ID NO:606	Cas3	N terminal
SEQ ID NO:607	S1Aspergillus	N terminal, C terminal

[00714] B. Cascade subunit protein fusion with another Cascade subunit protein

[00715] The two Cse2 proteins of the Cascade complex were fused together using a structure-guided approach based on the *E. coli* Cascade crystal structure 5H9E.pdb (www.rcsb.org/pdb/; see, e.g., Hayes, R.P, et al., Nature 530(7591):499-503 (2016)). Briefly, the C-terminus of one Cse2 and the N-terminus of a second Cse2 were fused together using a 10-aa flexible linker (SEQ ID NO:589). The full sequence of the Cse2-Cse2 (CasB_CasB) fusion protein is shown in SEQ ID NO:588.

[00716] In silico designed DNA sequences were provided to a commercial manufacturer (GenScript, Piscataway, NJ) for synthesis. The DNA sequences were cloned into the expression vector designed in Example 2 (SEQ ID NO:441). The Cse2 sequence was exchanged with SEQ ID NO:588.

[00717] Each expression vector was transfected into *E. coli* BL21 Star™ (Thermo Fisher Scientific, Waltham, MA) cells with a second vector encoding a guide RNA for the J3 target (SEQ ID NO:444), as described in Example 2. The *E. coli* Type I-E Cascade complex containing Cas5, Cas6, Cas7, Cse2-Cse2, and Cas8 was expressed and purified as described in Example 4B and 5B. Purification of the Cascade complexes comprising the fused Cse2-Cse2 variant demonstrate that fused Type I-E CRISPR-Cas subunit proteins successfully formed Cascade complexes having essentially the same composition (based on molecular weight) as Cascade complexes comprising wild-type proteins.

[00718] C. Electrophoretic mobility shift assays (EMSA) of Cascade/Cse2-Cse2 and J3 target

[00719] Purified Cascade/Cse2-Cse2 complexes were purified as described in this Example and subjected to an EMSA to demonstrate specific binding to their respective target sequence. Briefly, Cascade/ Cse2-Cse2 and Cascade/WT-Cse2 were purified and concentrated to 10 mg/mL. Cy5 double-stranded target DNA was produced as described in Example 6A and diluted to 1M in TE buffer (J3 target SEQ ID NO:469 and SEQ ID NO:472 and CCR5 target SEQ ID NO:474 and SEQ ID NO:470). Cascade complexes and labeled double-stranded target DNA were incubated for 30 min at 37°C at different protein/target ratios. Immediately following the incubation, 2 µl of 50% glycerol was added to the samples and they were loaded on a 5% native PAA gel. Gels were run at 4°C at 70V for 90 min in 0.5xTBE buffer and imaged on an AZURE c600 Bioimager (Azure BioSystems, Dublin, CA) and the bands were quantitated. The data are presented in Table 39.

[00720]

Table 39
Results of Cascade/Cse2-Cse2 EMSA

Cascade ID and guide	Target DNA	Cascade:dsDNA ratio	Gel shift %
Cascade/wt-Cse2 J3	J3	6.7	44
Cascade/ Cse2-Cse2 J3	J3	6.7	46
Cascade/wt-Cse2 J3	CCR5	6.7	LOD*
Cascade/ Cse2-Cse2 J3	CCR5	6.7	LOD

*LOD = below the limit of detection

[00721] D. Cascade subunit protein fusion with another Cascade subunit protein and an enzymatic protein domain

[00722] The cytidine deaminase rAPOBEC1 (apolipoprotein B mRNA editing enzyme catalytic subunit 1, *Rattus norvegicus*; NCBI Gene ID: 25383, uEnsembl:ENSRNOG00000015411) was selected for fusion. The Cse2-Cse2 protein was fused with rAPOBEC1 using a structure-guided approach based on the *E. coli* Cascade crystal structure 5H9E.pdb (www.rcsb.org/pdb/; see, e.g., Hayes, R.P, et al., Nature 530(7591):499-503 (2016)). Briefly, the C-terminus of rAPOBEC1 (SEQ ID NO:590) was fused to the N-terminus of the Cse2-Cse2 dimer (described above) using a 9-aa flexible linker (SEQ ID NO:591). The full sequence of the rAPOBECI_Cse2-Cse2 fusion protein is shown in SEQ ID NO:592.

[00723] In silico designed DNA sequences were provided to a commercial manufacturer (GenScript, Piscataway, NJ) for synthesis. The DNA sequences were cloned into the expression vector (SEQ ID NO:441), replacing the Cse2 sequence. Each expression vector was transfected into *E. coli* BL21 Star™ (Thermo Fisher Scientific, Waltham, MA) cells with a second vector encoding a guide RNA for the J3 target (SEQ ID NO:444), as described in Example 2. The *E. coli* Type I-E Cascade complex containing Cas5, Cas6, Cas7, rAPOBEC1_Cse2-Cse2, and Cas8 was expressed and purified as described in Example 4B and 5B. Purification of the Cascade complexes comprising the fused rAPOBEC1_Cse2-Cse2 variant demonstrate that cytidine deaminase fusions to Type I-E CRISPR-Cas subunit proteins were successfully used to form Cascade complexes having essentially the same composition (based on molecular weight) as Cascade complexes comprising wild-type proteins. Table 40 presents examples of enzyme fusions with Cse2-Cse2.

[00724]

Table 40
Other Enzyme Fusions (Such as Cytidine Deaminases) with Cse2-Cse2

SEQ ID NO:	Enzyme	Fusion to Cse2-Cse2
SEQ ID NO:608	rAPOBEC1	N terminal
SEQ ID NO:609	AID	C terminal
SEQ ID NO:610	CPmCDA1	C terminal
SEQ ID NO:611	Human APOBEC1	N terminal
SEQ ID NO:612	Human APOBEC3F	N terminal
SEQ ID NO:613	APOBEC3G	N terminal

[00725]

Example 12

Cascade Subunit Protein Fusions to Transcription Activation/Repression Domains

[00726] This Example illustrates the design of a *E. coli* Type I-E cp-Cas7 protein fused to a VP64 activation domain to confer transcriptional activation activity to the Cascade complex.

[00727] VP64 is a transcriptional activator comprising four tandem copies of VP16 (herpes simplex viral protein 16, DALDDFDLDML (SEQ ID NO:614); amino acids 437-447, UNIPROT:UL48) connected with glycine-serine (GS) linkers. When fused to a protein domain

that can bind near the promoter of a gene, VP64 (SEQ ID No:615) acts as a strong transcriptional activator. The *E. coli* Type I-E cp-Cas7 V2 (SEQ ID NO:616) can be selected for engineering.

[00728] The activation domain VP64 can be fused to the N-terminus of cpCas7 V2 (described in Example 10A). A linker (e.g., 5 to 50 amino acids in length) can be selected to operably link cpCas7 V2 and the VP64 domain.

[00729] In silico designed DNA sequences can be provided to a commercial manufacturer for synthesis. The DNA sequences encoding a VP64-cpCas7 V2 fusion protein can be cloned into an expression vector (e.g., SEQ ID NO:455, wherein VP64-cpCas7 V2 can be used to replace Cas7). Each expression vector can be transfected into *E. coli* BL21 Star™ (Thermo Fisher Scientific, Waltham, MA) cells with a second vector encoding a guide RNA for the J3 target (SEQ ID NO:444), as described in Example 2. The *E. coli* Type I-E Cascade complex containing Cas5, Cas6, VP64_cpCas7 V2, Cse2, and Cas8 can be expressed and purified as described in Examples 4 and 5. Purification of the Cascade complexes comprising the fused VP64_cpCas7 V2 variant can be used to form Cascade complexes having essentially the same composition (based on molecular weight) as Cascade complexes comprising wild-type proteins.

[00730] Selection of a guide targeted to the promoter region of a particular gene can be used to verify the ability of the Cascade complex comprising the fused VP64_cpCas7 V2 to facilitate transcriptional activation of the gene.

[00731]

Example 13

Site-Directed Recruitment of Functional Domains Fused to Cascade Subunit by dCas9/Guide Complex

[00732] This Example describes a method of engineering a Class 2 Type II CRISPR sgRNA, crRNA, tracrRNA, or crRNA and tracrRNA sequence with a Class 1 Type I CRISPR repeat stem sequence (e.g., a Type I-F CRISPR repeat stem sequence) for the recruitment of one or more Cascade subunit proteins (i.e., Cas6, Cas5, etc.) fused to a functional domain, to a Type II CRISPR Cas protein/guide RNA complex binding site. This method here is adapted from Gilbert, L., et. al., Cell 154(2):442-451 (2013) and Ferry, Q, et. al., Nature Communication 8:14633 doi: 10.1038/ncomms14633 (2017).

[00733] A. Engineering a Type II Guide RNA

[00734] A Type II CRISPR sgRNA, crRNA, tracrRNA, or crRNA and tracrRNA (collectively referred to as "Type II guide RNA") can be selected for engineering.

[00735] A Type II guide RNA sequence can be evaluated in silico for regions of incorporation of a Type I CRISPR repeat stem sequence. The Type I CRISPR repeat stem sequence can be attached at the 5' or 3' end of the Type II guide RNA, internal to the Type II guide RNA, or can replace secondary structure in the Type II guide RNA (e.g., 3' hairpin elements). Incorporation of the Type I CRISPR repeat stem sequence can be accompanied by a linker element nucleotide sequence. An example of a Type II tracrRNA 3' engineered to comprise a Type I CRISPR repeat stem sequence is presented in Table 41.

[00736]

Table 41
Exemplary Type II tracrRNA with 3' Type I CRISPR Repeat Stem Sequence

SEQ ID NO:	Sequence*
SEQ ID NO:617	5'-AGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUAAGUUCAcugccguauaggcagCUUU– 3'

*Type I CRISPR repeat stem sequence is underlined and in lower case letters. A corresponding DNA coding sequence is presented as SEQ ID NO:618.

[00737] A mammalian gene, such as C-X-C chemokine receptor type 4 (CXCR4), can be selected for targeting. The junction between the 5' UTR and exon 1 can be scanned in silico for a Type II CRISPR Cas protein target sequence occurring adjacent a Type II CRISPR Cas protein PAM sequence (e.g., 5'-NGG). The 20-nucleotide target sequence occurring upstream, in a 5' direction, can be incorporated into the Type II crRNA. An example of a Type II crRNA targeting CXCR4 is shown in Table 42.

[00738]

Table 42
Exemplary Type II crRNA Targeting CXCR4

SEQ ID NO:	Sequence*
SEQ ID NO:619	5'-GAACCAGCGGUUACCAUGGAGUUUUAGAGCUAUGCU–3'

*A corresponding DNA coding sequence is presented as SEQ ID NO:620.

[00739] Alternatively, the 3' end of the CXCR4 targeting spacer (RNA) (SEQ ID NO:619) can be covalently linked to the 5' end of the Type II tracrRNA with 3' Type I CRISPR repeat stem sequence (RNA) (SEQ ID NO:617) with a linker. A suitable linker element is 5'-GAAA-3'.

[00740] In silico designed Type II guide RNAs with the incorporated Type I CRISPR repeat stem sequence can be provided to a commercial manufacturer for synthesis.

[00741] A Type I Cascade subunit protein (e.g., Cas6) can be operably linked to a transcriptional activation or repression domain (e.g., KRAB) and c-terminally tagged with a nuclear localization signal (NLS) as described in Example 12.

[00742] A Type II Cas protein (e.g., Cas9) can be mutated such that it is catalytically inactive (e.g. dCas9) and tagged with a NLS sequence.

[00743] The Cas6-KRAB-NLS protein and the dCas9-NLS protein can be recombinantly expressed and purified from *E.coli.*

[00744] RNP complexes can be formed at a concentration of 60pmol dCas9 protein:60pmol Cas6-KRAB-NLS:120pmol:CXCR4 targeting crRNA:120pmol tracrRNA 3' engineered to comprise a Type I CRISPR repeat stem sequence. Prior to assembly with the dCas9 and the Cas6-KRAB-NLS, each of the 120pmol CXCR4 targeting crRNA and 120pmol tracrRNA 3' engineered to comprise a Type I CRISPR repeat stem sequence (herein referred to as "engineered Type II guide RNA") can be diluted to the desired total concentration (120pmol) in a final volume of 2µL, incubated for 2 minutes at 95°C, removed from a thermocycler, and allowed to equilibrate to room temperature. dCas9 and the Cas6-KRAB-NLS protein can be diluted to an appropriate concentration in binding buffer (20mM HEPES, 100mM KCl, 5mM $MgCl_2$, and 5% glycerol at pH 7.4) to a final volume of 3µL and mixed with the 2µL of Type II guide RNA, followed by incubation at 37°C for 30 minutes. An un-transfected control (e.g., buffer only), un-engineered Type II guide RNA, or a Cas6 not linked to a repression domain, can be used to assemble negative control RNPs.

[00745] B. Cell transfections using dCas9:Cas6-KRAB-NLS: engineered Type II guide RNA

[00746] dCas9:Cas6-KRAB-NLS: engineered Type II guide RNA nucleoprotein complexes can be transfected into HEK293 cells (ATCC, Manassas VA), using the Nucleofector® 96-well Shuttle System (Lonza, Allendale, NJ) and the following protocol: The complexes can be dispensed in a 5µL final volume into individual wells of a 96-well plate. The cell culture medium can be removed from the HEK293 cell culture plate and the cells detached with TrypLE™ (Thermo Scientific, Wilmington, DE). Suspended HEK293 cells can be pelleted by centrifugation for 3 minutes at 200 x g, TrypLE reagents aspirated, and cells washed with calcium and magnesium-free phosphate buffered saline (PBS). Cells can be pelleted by centrifugation for 3 minutes at 200 x g, the PBS aspirated, and the cell pellet re-suspended in 10mL of calcium and magnesium-free PBS.

[00747] The cells can be counted using the Countess® II Automated Cell Counter (Life Technologies; Grand Island, NY). 2.2 x 10^7 cells can be transferred to a 1.5ml microfuge tube

and pelleted. The PBS can be aspirated and the cells re-suspended in Nucleofector™ SF (Lonza, Allendale, NJ) solution to a density of 1 x 10^7 cells/m. 20µL of the cell suspension can be then added to each individual well containing 5µL of RNP complexes, and the entire volume from each well can be transferred to a well of a 96-well Nucleocuvette™ (Lonza, Allendale, NJ) plate. The plate can be loaded onto the Nucleofector™ 96-well Shuttle™ (Lonza, Allendale, NJ) and cells nucleofected using the 96-CM-130 Nucleofector™ (Lonza, Allendale, NJ) program. Post-nucleofection, 70µL Dulbecco's Modified Eagle Medium (DMEM; Thermo Scientific, Wilmington, DE), supplemented with 10% Fetal Bovine Serum (FBS; Thermo Scientific, Wilmington, DE), penicillin and streptomycin (Life Technologies, Grand Island, NY) can be added to each well, and 50µL of the cell suspension can be transferred to a 96-well cell culture plate containing 150µL pre-warmed DMEM complete culture medium. The plate can be transferred to a tissue culture incubator and maintained at 37°C in 5% CO_2 for 48 hours.

[00748] 72 hours after nucleofection of the dCas9:Cas6-KRAB-NLS: engineered Type II guide RNA nucleoprotein complexes, cells can be evaluated for repression of CXCR4 expression. Culture medium can be aspirated from the HEK293, and the cells can be washed once with calcium and magnesium-free PBS then are trypsinized by the addition of TrypLE (Life Technologies, Grand Island, NY) followed by incubation at 37ºC for 3-5 minutes. Trypsinized cells can be gently pipetted up and down to form a single cell suspension, and the cells can then be pelleted by centrifugation for 3 minutes at 200 x g. After centrifugation, the culture medium can be aspirated and cells are re-suspended in a 10mM EDTA/PBS buffer and gently mixed into a singles cell suspension. The single-cell suspension can be stained using 0.05% FITC conjugated to an anti-human CXCR4 antibodies (Medical & Biological Laboratories Co., Nagoya, Japan) in PBS containing 10% FBS for 1 hour at room temperature. Isotype controls and native RNP controls can be similarly stained for reference. Stained cells can then be sorted LSR II flow cytometer (BD laboratories, San Jose, CA) and population of FITC positive fluorescent cells tallied.

[00749] Reduction in CXCR4 expression is measure by a decrease in detected fluorescence of a dCas9:Cas6-KRAB-NLS: engineered Type II guide RNA nucleofected sample compared to the measured fluorescence of a non-transfected control. Decrease in fluorescence from the flow cytometer can be used to demonstrate that a engineered Type II guide RNA with a Type I CRISPR repeat stem sequence can be used in combination with a nuclease-deficient Type II Cas9 protein to recruit and localize a Type I CRISPR Cascade subunit protein fused to repression domain to a gene target and repress transcription of said gene target.

[00750]

Example 14

Identification and Screening of Type I *cas* Genes

[00751] This Example describes a method to identify and screen Type I *cas* genes from different species. The method presented here is adapted from Shmakov, S., et al., Mol. Cell 60:385-397 (2015).

[00752] A. Identification of Type I CRISPR-Cas genes

[00753] Using the Basic Local Alignment Search Tool (BLAST, blast.ncbi.nlm.nih.gov/Blast.cgi), a search of the genomes of various species can be conducted to identify one or more genes coding for the various gene component of the Type I CRISPR-Cas complex. The *cas1* integrase gene is a component of both Class 1 and Class 2 CRISPR-Cas families, and upon identification of species containing the *cas1* gene, subsequence searcher in these genomes can be conducted to isolate genomes comprising Type I-specific genes. Genome searches can be anchored upon the CRISPR-Cas integrase genes *cas1,* an exemplary *cas1* sequence from the Type I-E system from *E. coli* K-12 MG1655 that can be used is SEQ ID. NO:621. Particular genes (e.g., *cas7* and *cas5*) are core components of the interference complexes of the Type I systems and can be used to further differentiate species containing Type I systems. Exemplary sequences of *E. coli* K-12 MG1655 *cas7* and *cas5* genes that can be used are SEQ ID. NO:622 and SEQ ID. NO:623, respectively. Genomes identified possessing *cas7* and *cas5* genes can be further parsed through the identification of the Type I-specific nuclease-helicase *cas3* gene or homologs thereof. An exemplary sequence of a *E. coli* K-12 MG1655 *cas3* sequence that can be used is SEQ ID. NO:624.

[00754] Genomes containing CRISPR-Cas integrase genes *cas1*, Type I interference complex genes *cas7* and *cas5,* and the nuclease-helicase *cas3* gene, or some combination thereof, are likely candidates of Type I CRISPR-Cas system(s). Type I CRISPR-Cas genes are generally found in proximity to one in a single genomic locus, typically within 20 kilobases (kb). The area around the *cas1*, *cas7*, *cas5*, or *cas3* genes can be searched for other open reading frames (ORFs) of the remaining *cas* genes that constitute a Type I interference complex. The amino acid sequence of putative ORFs can be compared to known Type I genes for homology or the presence of characteristic protein domains of the Type I protein components can be analyzed using the homology detection and structure prediction search tools available through the Max Planck Institute Bioinformatics Toolkit (www.toolkit.tuebingen.mpg.de/#/), or equivalent.

[00755] B. Screening of identified Type I components

[00756] Once a putative collection of Type I components (e.g., *cas* genes and the corresponding crRNA) have been identified, the Type I components can be tested for their ability to carry out programmable DNA targeting.

[00757] Putative *cas* genes and the crRNA can be encoded into expression vectors following the guidance of Examples 1, 2, and 3. Vectors encoding the various *cas* genes and crRNA can be introduced into a bacteria strain and the Type I interference complex expressed and purified as described in Examples 4 and 5. The elution fraction from the size –exclusion chromatography (SEC) column, can be analyzed via SDS-PAGE gel to determine the identity, based on weight, of the protein components comprising a complete Type I interference complex. An ethidium bromide gel can also be run to detect the presence of crRNA as part of the interference complex.

[00758] Purified Cascade complexes can be tested for their ability to support *in vitro* biochemical cleavage of a DNA target as described in Examples 6 and 7.

[00759] Control expressions and purification samples, where single putative *cas* gene are not expressed, can be used to determine the required *cas* genes that constitute a complete Type I interference complex capable of programmable DNA target.

[00760] For certain applications, identification of individual *cas* gene homologs (e.g., *cas7*) from a genomic sequence is sufficient and additional *cas* genes need not be identified nor screening performed.

[00761]

Example 15

Identification of Type I crRNAs

[00762] This Example describes a method to identify Type I crRNAs in different species. The method presented here is adapted from Chylinski, K., et al., RNA Biology 10:726-737 (2013).

[00763] A search of genomes of various species can be conducted to identify Type I CRISPR-Cas genes as described in Example 17A. Genomes that comprise one of more Type I specific *cas* genes are candidate genomes that likely to contain CRISPR RNAs (crRNAs) encoded within the CRISPR repeat-spacer array. The sequences adjacent to the identified Type I *cas* genes (e.g., a *cas7*, *cas5*, or *cas3* gene) can be probed for an associated CRISPR repeat-spacer array. Methods for in silico predictive screening can be used to extract the crRNA sequence from the repeat array following Grissa, I.V., et. al. Nucleic Acids Res. 35(Web Server issue):W52-W57 (2007). The crRNA sequence is contained within CRISPR repeat array and can be identified by its hallmark repeating sequences interspaced by foreign spacer sequences.

[00764] A. Preparation of RNA-seq library

[00765] The putative CRISPR array containing the individual crRNA identified in silico can be further validated using RNA sequencing (RNA-seq).

[00766] Cells from species identified as comprising putative Type I *cas* genes and crRNA components can be procured from a commercial repository (e.g., ATCC, Manassas, VA; German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), Braunschweig, Germany).

[00767] Cells can be grown to mid-log phase and total RNA prepped using Trizol reagent (SigmaAldrich, St. Louis, MO) and treated with DNaseI (Fermentas, Vilnius, Lithuania).

[00768] 10μg of the total RNA can be treated with Ribo-Zero rRNA Removal Kit (Illumina, San Diego, CA) and the remaining RNA purified using RNA Clean and Concentrators (Zymo Research, Irvine, CA).

[00769] A library can be prepared using a TRUSEQ™ Small RNA Library Preparation Kit (Illumina, San Diego, CA), following the manufacturer's instructions. This will result in cDNAs having adapter sequences.

[00770] The resulting cDNA library can be sequenced using MiSeq Sequencer (Illumina, San Diego, CA).

[00771] B. Processing of sequencing data

[00772] Sequencing reads of the cDNA library can be processed, for example, using the following method.

[00773] Adapter sequences can be removed using cutadapt 1.1 (pypi.python.org/ pypi/cutadapt/1.1) and about 15 nucleotides trimmed from the 3'end of the read to improve read quality.

[00774] Reads can be aligned to the genome of the respective species (i.e., from which the putative crRNA is to be identified) using Bowtie 2 (www.bowtie-bio.sourceforge.net/bowtie2/index.shtml). The Sequence Alignment/Map (SAM) file, which is generated by Bowtie 2, can be converted into a Binary Alignment/Map (BAM) file using SAMTools (www.samtools.sourceforge.net/) for subsequent sequencing analysis steps.

[00775] Read coverage mapping to the CRISPR locus or loci can be calculated from the BAM file using BedTools (bedtools.readthedocs.org/ en/latest/).

[00776] The BED file, as generated in the previous step, can be loaded into Integrative Genomics Viewer (IGV; www.broadinstitute.org/igv/) to visualize the sequencing read pileup. Read pile can be used to identify the 5' and 3' termini of the transcribed putative crRNA

sequence. The RNA-seq data can be used to validate that a putative crRNA element is actively transcribed *in vivo*.

[00777] Putative crRNA can be tested with their cognate Type I *cas* genes for the ability to carry out programmable DNA targeting following the guidance of Example 17A.

[00778]

Example 16

Probing for Sites Tolerant of Changes in Cascade Guide RNA Backbones

[00779] This Example describes the generation and testing of various changes to Type I guide crRNAs and their suitability for use in constructing Cascade polynucleotide complexes. The method described below is adapted from Briner, A., et al., Mol. Cell 56:333–339 (2014).

[00780] Changes can be introduced into the crRNA backbone, and the resulting engineered crRNA tested with a cognate Cascade complex to facilitate the identification of regions or positions in the Type I guide crRNA backbone amenable to engineering.

[00781] A crRNA from a Type I CRISPR system (e.g., *E. coli* Cascade) can be selected for engineering. The crRNA sequence can be engineered in silico to introduce one or more base changes (e.g., substitutions, changes, mutations, deletions, and/or insertions into nucleic acid sequences in regions selected from one or more of the following regions: nucleic acid sequences 5' of the spacer (5' handle), the spacer element, Type I CRISPR repeat stem sequence, or 3' of the Type I CRISPR repeat stem sequence (3' handle).

[00782] Base changes can also be used to introduce mismatches in the hydrogen base-pair interactions of any of the crRNA regions, or base-pair mutation introducing an alternative hydrogen base-pair interaction through substitution of two bases, wherein the alternative hydrogen base-pair interaction differs from the original hydrogen base-pair interaction (e.g., the original hydrogen base-pair interaction is Watson-Crick base pairing and the substitution of the two bases form a reverse Hoogsteen base pairing). Substitution of bases can also be used to introduce hydrogen base-pair interaction within the crRNA backbone.

[00783] Regions of the crRNA can be independently engineered to introduce secondary structure elements into the crRNA backbone. Such secondary structure elements include, but are not limited to, the following: stem-loop elements, stem elements, pseudo-knots, and ribozymes. Furthermore, the crRNA backbone can be engineered to delete portions of the crRNA backbone either through deletion at the 5' end, 3' end, or internal to the crRNA. Alternative backbone structures can also be introduced.

[00784] In silico designed crRNA sequences can be provided to a commercial manufacturer for synthesis.

[00785] Engineered crRNAs can be evaluated for their ability to support binding by individual Cascade subunit proteins (i.e., Cas6, Cas5, etc.), or to support complete formation of the Cascade protein complex, or to support formation of the Cascade complex and modification of a double-stranded DNA target sequence through recruitment of a nuclease (e.g., Cas3). crRNA binding to individual Cascade subunit proteins and Cascade protein complex assembly can be evaluated by nano-ESI mass spectrometry in a manner similar to Jore, M., et al., Nature Structural & Molecular Biology 18:529–536 (2011). Biochemical characterization of crRNA and Cascade protein complex modification of a double-stranded DNA target sequence through recruitment of a nuclease can be carried out in a manner similar to those described in Examples 6 and 7. Engineered crRNA that are capable of supporting formation of the Cascade complex and modification of a double-stranded DNA target sequence through recruitment of a nuclease can be validated for activity in cells using the method described in Example 8A, Example 8B, Example 8C, and Example 8D.

[00786]

Example 17

Screening of Cascade Complex Guides Comprising DNA Target Binding Sequences

[00787] This Example illustrates the use of Type I CRISPR proteins and Type I guide crRNAs of the present invention to modify DNA target sequences present in human gDNA (gDNA) and to measure the level of cleavage activity at those sites.

[00788] Target sites (DNA target sequences) can be first selected from gDNA. Type I guide crRNAs can be designed to target the selected sequences. Assays (e.g., as described in Example 7) can be performed to determine the level of DNA target sequence cleavage.

[00789] A. Selecting DNA target sequences from gDNA

[00790] PAM sequences (e.g., ATG) for a Cascade protein complex (e.g., *E. coli* Type I-E Cascade) can be identified within the selected genomic region.

[00791] One or more Cascade DNA target sequences (e.g., 32 nucleotides in length) that are 3' adjacent to an ATG PAM sequence can be identified.

[00792] Criteria for selection of nucleic acid target sequences can include, but are not limited to, the following: homology to other regions in the genome; percent G-C content; melting temperature; presences of homopolymer within the spacer; distance between the two sequences; and other criteria known to one skilled in the art.

[00793] A DNA target binding sequence that hybridizes to the Cascade DNA target sequence can be incorporated into a guide crRNAs. The nucleic acid sequence of a guide crRNA construct is typically provided to and synthesized by a commercial manufacturer.

[00794] A guide crRNA, as described herein, can be used with cognate Type I Cascade protein complex to form crRNA/Cascade protein complexes.

[00795] B. Determination of cleavage percentages and specificity

[00796] *In vitro* cleavage percentages and specificity (i.e., the amount of off-target binding) related to a guide crRNA can be determined, for example, using the cleavage assays described in Example 7, and compared as follows:

[00797] (1) If only a single DNA target sequences is identified or selected for a guide crRNA, the cleavage percentage and specificity for each of the DNA target sequences can be determined. If so desired, cleavage percentage and/or specificity can be altered in further experiments using methods including, but not limited to, engineering the guide crRNA, or introducing effector proteins/effector protein-binding sequences to engineer the guide crRNA or the Cascade subunit proteins, or ligand/ligand-binding moieties to engineer the guide crRNA or the Cascade subunit proteins.

[00798] (2) If multiple DNA target sequences are identified or selected for guide crRNAs, the percentage cleavage data and site-specificity data obtained from the cleavage assays can be compared between different DNAs comprising the target binding sequence to identify the DNA target sequences having the desired cleavage percentage and specificity. Cleavage percentage data and specificity data provide criteria on which to base choices for a variety of applications. For example, in some situations the activity of the guide crRNA may be the most important factor. In other situations, the specificity of the cleavage site may be relatively more important than the cleavage percentage. If so desired, cleavage percentage and/or specificity can be altered in further experiments using methods including, but not limited to, engineering the guide crRNA, introducing effector proteins/effector protein-binding sequences to engineer the guide crRNA or the Cascade subunit proteins, or ligand/ligand-binding moieties to engineer the guide crRNA or the Cascade subunit proteins.

[00799] Alternatively, or in addition to the *in vitro* analysis, in-cell cleavage percentages and specificities of guide crRNAs can be obtained using, for example, the method described in Example 8C and Example 8D, and compared as follows:

[00800] (1) If only a single DNA target sequences is identified or selected for a guide crRNA, the cleavage percentage and specificity for each of the DNA target sequences can be determined. If so desired, cleavage percentage and/or specificity can be altered in further experiments using methods including, but not limited to, engineering the guide crRNA, or introducing effector proteins/effector protein-binding sequences to engineer the guide crRNA

or the Cascade subunit proteins, or ligand/ligand-binding moieties to engineer the guide crRNA or the Cascade subunit proteins.

[00801] (2) If multiple DNA target sequences are identified or selected for guide crRNAs, the percentage cleavage data and site-specificity data obtained from the cleavage assays can be compared between different DNAs comprising the target binding sequence to identify the DNA target sequences having the desired cleavage percentage and specificity. Cleavage percentage data and specificity data provide criteria on which to base choices for a variety of applications. For example, in some situations the activity of the guide crRNA may be the most important factor. In other situations, the specificity of the cleavage site may be relatively more important than the cleavage percentage. If so desired, cleavage percentage and/or specificity can be altered in further experiments using methods including, but not limited to, engineering the guide crRNA, introducing effector proteins/effector protein-binding sequences to engineer the guide crRNA or the Cascade subunit proteins, or ligand/ligand-binding moieties to engineer the guide crRNA or the Cascade subunit proteins.

[00802]

Example 18

Varying FokI-Cas8 Linker Composition and Interspacer Distance for Efficient FokI-Cascade Complex Genome Editing

[00803] This Example illustrates the design and testing of multiple fusion proteins comprising FokI-Cas8 and linker polypeptides of various lengths, as well as the effect of varying interspacer distances for efficient genome editing.

[00804] A. Production of a vector encoding *E. coli* Type I-E Cascade complex components comprising FokI fusion proteins to be transfected into target cells

[00805] Minimal CRISPR arrays were designed to target a set of loci in the human genome at or near two different genes: *ADAMTSL1* and *PCSK9*. Interspacer distances ranged from 14-60 bp, in increments of 2 bp. Four targets were designed for each interspacer distance. Targets were flanked by either AAG or ATG PAM sequences. Coding sequences for guides containing "repeat-spacer-repeat-spacer-repeat" sequences were cloned as described in Example 9A with SEQ ID NO:454. SEQ ID NO:625 through SEQ ID NO:816 provide the sequences for the full set of oligonucleotide sequences used to generate the minimal CRISPR arrays.

[00806] FokI-Cascade RNP subunit protein component-encoding genes were cloned into vectors comprising: CMV promoters to enable delivery and expression in mammalian cells; *cas* genes linked via 2A viral peptide "ribosome-skipping" sequences; a fusion protein comprising

FokI and Cas8 connected with a 30-aa linker (SEQ ID NO:455). Additional linker polypeptide sequences of varying length and amino acid composition were designed and used to connect FokI to the Cas8 protein in these vectors. The additional linker polypeptide sequences are listed in Table 43.

[00807]

Table 43
Amino Acid Linker Sequences

SEQ ID NO:	Linker length (amino acids)	Amino acid sequence
SEQ ID NO:817	5	GGGGS
SEQ ID NO:818	8	TGPGAAAR
SEQ ID NO:819	10	GGSGSSGGSG
SEQ ID NO:820	15	GGSGSSGGSGSSGGS
SEQ ID NO:821	17	ADPTNRAKGLEAVSVAS
SEQ ID NO:822	20	SGSETPGTSESATPESGGSG
SEQ ID NO:433	30	SGSETPGTSESATPESGGSGSSGGSGSSGG
SEQ ID NO:823	40	SGSETPGTSESATPESGGSGSSGGSGSSGGSGS SGGSGSS
SEQ ID NO:824	50	SGSETPGTSESATPESGGSGSSGGSGSSGGSGS SGGSGSSGGSGSSGGSG

[00808] B. Transfection of vectors encoding FokI-Cascade RNP complex components

[00809] Transfection conditions were essentially as described in Example 8B with the following modifications. Prior to nucleofection, 5 µl of plasmid vector solution was transferred to individual wells of a 96-well plate. Each well contained 2.4 µg of plasmid encoding FokI-Cascade RNP complex subunit protein components and ~1-2 µg of plasmid encoding the minimal CRISPR array.

[00810] C. Deep sequencing of gDNA from transfected cells

[00811] Deep sequencing was performed essentially as described in Example 8C with the following modifications. Instead of primers Y and Z from Table 36 of Example 8C, the target-specific primers were SEQ ID NO:825 to SEQ ID NO:1016.

[00812] D. Deep sequencing data analysis

[00813] Deep sequencing data analysis was performed essentially as described in Example 8D. FIG. 31A and FIG. 31B present the results of the data analysis. In FIG. 31A and FIG. 31B, percent genome editing is shown as a function of FokI-Cas8 linker type (FIG. 31A, FIG. 31B, horizontal axis 5-50 AA) and interspacer distance (n=1) (FIG. 31A, FIG. 31B, vertical axis, 14-60 bp. In FIG. 31A, the grey scale vertical bar to the right is percentage of

indels. In FIG. 31B, values in the cells are the percent indels. An initial analysis of the data showed genome editing was highest with FokI-Cas8 linkers of 17 and 20 amino acids (SEQ ID NO:821 and SEQ ID NO:822, respectively) and with interspacer distances of ~26bp and ~30-32bp. The data was reprocessed and samples with less than a thousand sequences reads were removed as they may contain inflated editing values due to low coverage (sites were only retained if all the associated samples contained >1000 reads). This data, presented in FIG. 31A and FIG. 31B, showed that genome editing was highest with FokI-Cas8 linkers of 17 and 20 amino acids (SEQ ID NO:821 and SEQ ID NO:822, respectively) and with interspacer distances of ~30-32bp. Thus, efficient genome editing using Type I CRISPR-Cas complexes comprising Fok1-Cas8 fusion proteins was achieved by varying the interspacer distance and the linker polypeptide length of the FokI-Cas8 fusion protein. The amino acid composition of the linker polypeptides is discussed herein.

[00814]

Example 19

Identifying Cascade Homologs for Genome Editing

[00815] This Example illustrates the design and testing of multiple homolog Cascade complexes to evaluate the efficiency of genome editing.

[00816] A. Identification of sites for testing with homolog Cascade complexes

[00817] A panel of sites was identified for testing additional homolog Cascade complexes. Specifically, minimal CRISPR arrays were designed to target a set of loci in the human genome with 30-bp interspacer distances and that were flanked by either AAG or ATG PAM sequences. Guide polynucleotides containing “repeat-spacer-repeat-spacer-repeat” sequences were cloned following the method described in Example 9A with SEQ ID NO:454. The full set of oligonucleotide sequences used to generate the minimal CRISPR arrays are presented as SEQ ID NO:1017 to SEQ ID NO:1130 (Hsa33F, SEQ ID NO:1017, and Hsa33R, SEQ ID NO:1074, exemplify one pair). A positive control comprising guides targeting the TRAC locus was included (SEQ ID NO:454).

[00818] FokI-Cascade RNP subunit protein component-encoding genes were cloned into vectors comprising: CMV promoters to enable delivery and expression in mammalian cells; *cas* genes linked via 2A viral peptide “ribosome-skipping” sequences; a fusion protein comprising FokI and Cas8 connected with a 30-aa linker (SEQ ID NO:455).

[00819] B. Transfection of vectors encoding FokI-Cascade RNP complex components

[00820] Transfection conditions were performed essentially as described in Example 8B with the following modifications. Prior to nucleofection, 5 µl of plasmid vector solution was transferred to individual wells of a 96-well plate. Each well contained 3 µg of plasmid encoding FokI-Cascade RNP subunit protein components and 0.3 µg of plasmid encoding the minimal CRISPR array.

[00821] C. Deep sequencing of gDNA from transfected cells

[00822] Deep sequencing was performed essentially as described in Example 8C with the following modifications. Instead of primers Y and Z from Table 36 of Example 8C, the target-specific primers used in this Example were SEQ ID NO:1131 to SEQ ID NO:1244.

[00823] D. Deep sequencing data analysis

[00824] Deep sequencing data analysis was performed essentially as described in Example 8D. FIG. 32 present the results of the data analysis. In FIG. 32, percent genome editing (FIG. 32, vertical axis, % editing) is plotted against 58 test sites (FIG. 32, horizontal axis, "Target"; oligonucleotide sequences used to generate these minimal CRISPR arrays are discussed above) in addition to target Hsa07 from Example 8A (n=3). As is shown in FIG. 32, editing ranged from ~6% to below the limit of detection. From these data, a panel of eight sites (Hsa07 as well as targets 1, 3-5, 10, 13, and 16 corresponding to the following targets Hsa37, Hsa43, Hsa46, Hsa60, Hsa77, Hsa88, and Hsa126) with AAG PAMs were selected for testing homolog Cascade complexes for genome editing.

[00825] E. Identification of homolog Cascade complexes to test with FokI nuclease for genome editing

[00826] Cas8 protein sequences from different Type I systems were used as queries for psi-BLASTp to generate phylogenetic trees for homolog selection. Specifically, Cas8 from *Fusobacterium nucleatum* (WP_008798978.1) was used for Type I-B, Cas8 from *Bacillus halodurans* (WP_010896519.1) was used for Type I-C, Cas8 from *E. coli* (WP_001050401.1) was used for Type I-E, Cas8 from *Pseudomonas aeruginosa* (WP_003139224.1) was used for Type I-F, and Cas5 from *Shewanella putrefaciens* (WP_011919226.1) was used for Type I-Fv2.

[00827] Next, psi-BLASTp was iterated multiple times until thousands of homologs were identified for each Type I system. From this information, phylogenetic trees were built using the interactive Tree of Life online software (iTOL, accessible at itol.embl.de/login.cgi). The trees were visually inspected after auto-collapsing clades using variable branch lengths.

[00828] Lists of organisms falling within major clades were then outputted and manually inspected for selection. In this step, priority was placed on selecting homologs that sampled from different regions of the phylogenetic tree, both for the 12 homologs within the Type I-E as well as 2-3 representative homologs for Types I-B, I-C, I-F, and I-Fv2. *cas8* and *cas5* candidates, based on the above phylogenetic analysis, were inputted into NCBI, and the genomic context within the endogenous host bacterium was visually inspected within NCBI's genome graphics browser. Cascade homologs were selected only if (1) they were found in organisms that grow at 37°C; (2) their *cas* gene operons were intact and had all the expected Cascade subunit protein encoding genes, a *cas3* gene, and intact acquisition genes (i.e., *cas1* and *cas2*); (3) their *cas* gene operon was flanked by one or more CRISPR arrays; and (4) their CRISPR arrays contained > 10 spacers. For some homologs, the CRISPRfinder program (crispr.i2bc.paris-saclay.fr/Server/) was used to identify putative PAM sequences. Based on the above criteria, the 22 homolog Cascade complexes shown in Table 44 were selected.

[00829]

Table 44
Homolog Cascade Complexes

SEQ ID NO:	Cascade homolog organism	PAM	Spacer length	Type
SEQ ID NO:1245	*Oceanicola sp.* HL-35	AAG	32	I-E
SEQ ID NO:1246	*Pseudomonas sp.* S-6-2	AAG	32	I-E
SEQ ID NO:1247	*Salmonella enterica subsp. enterica serovar* Muenster strain	AAG	32	I-E
SEQ ID NO:1248	*Atlantibacter hermannii* NBRC 105704	AAG	32	I-E
SEQ ID NO:1249	*Geothermobacter sp.* EPR-M	AAG	32	I-E
SEQ ID NO:1250	*Methylocaldum sp.* 14B	AAG	32	I-E
SEQ ID NO:1251	*Methanocella arvoryzae* MRE50	AAG	32	I-E
SEQ ID NO:1252	*Pseudomonas aeruginosa* DHS01	AAG	32	I-E
SEQ ID NO:1253	*Lachnospiraceae bacterium* KH1T2	GAA	35	I-E
SEQ ID NO:1254	*Klebsiella pneumoniae* strain VRCO0172	GAA	33	I-E
SEQ ID NO:1255	*Streptococcus thermophilus* strain ND07	GAA*	33	I-E
SEQ ID NO:1256	*Streptomyces sp.* S4	GAA	33	I-E

Table 44 Homolog Cascade Complexes				
SEQ ID NO:	**Cascade homolog organism**	**PAM**	**Spacer length**	**Type**
SEQ ID NO:1257	*Campylobacter fetus subsp. testudinum* Sp3	TCA	36	I–B
SEQ ID NO:1258	*Odoribacter splanchnicus* DSM 20712	TCA	36	I–B
SEQ ID NO:1259	*Bacillus halodurans* C-125	TTC	34	I–C
SEQ ID NO:1260	*Desulfovibrio vulgaris* RCH1 plasmid pDEVAL01	TTC	34	I–C
SEQ ID NO:1261	*Geobacillus thermocatenulatus* strain KCTC 3921	TTC	35	I–C
SEQ ID NO:1262	*Vibrio cholerae* strain L15 L15_contig8	CC	32	I–F
SEQ ID NO:1263	*Pseudomonas aeruginosa* UCBPP-PA14	CC	32	I–F
SEQ ID NO:1264	*Shewanella putrefaciens* CN-32	CC	32	I–Fv2
SEQ ID NO:1265	*Acinetobacter sp.* 869535	CC	32	I–Fv2
SEQ ID NO:1266	*Vibrio cholerae* HE48 vcoHE48.contig.11	CC	32	I–Fv2

* as identified by Sinkunas, T., et al., EMBO J. 32:385-394 (2013); however, the data presented herein demonstrates that *Streptococcus thermophilus* strain ND07 can utilize a single A as a PAM sequence *in vivo*.

[00830] F. Production of vectors encoding FokI-Cascade RNP components from 22 distinct species for transfection into target cells

[00831] Sequences for each *cas* gene from each homolog were synthesized as part of a polycistronic construct that included a fusion protein comprising FokI nuclease and Cas8. For each Type I-E Cascade complex homolog, a set of ~7-8 guides targeting loci with the appropriate PAM sequences were generated. For each Type I-B, I-C, I-F, and I-Fv2 Cascade homolog, a set of ~2-7 guides targeting loci with appropriate PAM sequences were generated. Each Cascade complex homolog system required unique repeat sequences to process their cognate guide (SEQ ID NO:1267 to SEQ ID NO:1288). Coding sequences for guides containing "repeat-spacer-repeat-spacer-repeat" sequences were cloned using the method described in Example 9A for SEQ ID NO:454. Oligonucleotides were phosphorylated on the 5' end and appended with overhang sequences to enable cloning into plasmid vectors with the appropriate repeat sequences. The full set of oligonucleotide sequences used to generate the minimal CRISPR arrays for the 22 Cascade complex homologs are presented as (SEQ ID NO:1289 to SEQ ID NO:1400).

[00832] FokI-Cascade RNP subunit protein component-encoding genes were cloned into vectors comprising: CMV promoters to enable delivery and expression in mammalian cells; *cas* genes linked via 2A viral peptide "ribosome-skipping" sequences; a fusion protein comprising FokI and Cas8 connected with a 30-aa linker.

[00833] G. Transfection of plasmids encoding FokI-Cascade RNP complexes

[00834] Transfection conditions were essentially as described in Example 8B with the following modifications. Prior to nucleofection, 5 µl of plasmid vector solution was transferred to individual wells of a 96-well plate. Each well contained 1.5 µg of plasmid encoding FokI-Cascade RNP subunit protein components and ~0.5-1.5 µg of plasmid encoding the minimal CRISPR array. Experiments were performed in triplicate and included FokI-Cascade RNP complexes from *E. coli* (SEQ ID NO:455) targeted to eight sites (Hsa07 from Example 8A and Hsa37, Hsa43, Hsa46, Hsa60, Hsa77, Hsa88, Hsa126 from Example 19F and Example 19G) as positive controls. As previously described, the following oligonucleotides were used to generate the minimal CRISPR arrays used with the *E. coli* positive control: Hsa37 (SEQ ID NO:1019; SEQ ID NO:1076), Hsa43 (SEQ ID NO:1024; SEQ ID NO:1081), Hsa46 (SEQ ID NO:1027; SEQ ID NO:1084), Hsa60 (SEQ ID NO:1037; SEQ ID NO:1094), Hsa77 (SEQ ID NO:1045; SEQ ID NO:1102), Hsa88 (SEQ ID NO:1050; SEQ ID NO:1107), Hsa126(SEQ ID NO:1072; SEQ ID NO:1129).

[00835] H. Deep sequencing of gDNA from transfected cells

[00836] Deep sequencing was performed essentially as described in Example 8C with the following modifications. Instead of primers Y and Z from Table 36 of Example 8C, the target-specific primers used in this Example were SEQ ID NO:1401 to SEQ ID NO:1512. For both Type I-E RNP complexes and Type I-B, I-C, I-F, and I-Fv2 RNP complexes, control samples comprising *E. coli* Type I-E Cascade were included for comparison and sequenced with target-specific primers corresponding to targets Hsa07 from Example 8A and Hsa37, Hsa43, Hsa46, Hsa60, Hsa77, Hsa88, Hsa126 from this Example. More specifically, the following target-specific amplification primers were used for these targets: Hsa37 (SEQ ID NO:1133; SEQ ID NO:1190), Hsa43 (SEQ ID NO:1138; SEQ ID NO:1195), Hsa46 (SEQ ID NO:1141; SEQ ID NO:1198), Hsa60 (SEQ ID NO:1151; SEQ ID NO:1208), Hsa77 (SEQ ID NO:1159; SEQ ID NO:1216), Hsa88 (SEQ ID NO:1164; SEQ ID NO:1221), Hsa126(SEQ ID NO:1186; SEQ ID NO:1243).

[00837] I. Deep sequencing data analysis

[00838] Deep sequencing data analysis was performed essentially as described in Example 8D. FIG. 33A and FIG. 33B show results from these experiments. In FIG. 33A, the

vertical axis is percent editing (FIG. 33A, % editing) and the numbers on the horizontal axis are SEQ ID Nos. corresponding to Type I-E homolog systems. Editing was observed with many of the Type I-E FokI-Cascade homologs (FIG. 33A). The highest editing was observed with the variant from *Pseudomonas sp. S-6-2*, while other homologs (i.e., *Salmonella enterica*, *Geothermobacter sp.* EPR-M, *Methanocella arvoryzae MRE50*, and *S. thermophilus* (strain ND07)) showed editing approximately equivalent to *E. coli*. In FIG. 33B, the vertical axis is percent editing (FIG. 33B, % editing) and the numbers on the horizontal axis are SEQ ID Nos. corresponding to Type I-B, I-C, I-F, and I-Fv2 homolog systems. Editing with FokI-Cascade RNPs derived from Types I-B, I-C, I-F, and I-Fv2 was below the limit of detection (FIG. 33B).

[00839] This Example provides methods for screening Type I homologs to identify Type I systems that provide genomic editing capability. Additional Type I homolog screening is described in Example 22.

[00840]

Example 20

Varying FokI-Cas8 Linker Length and Interspacer Distances in *Pseudomonas sp S-6-2* for Efficient Genome Editing

[00841] This Example illustrates the design and testing of multiple fusion proteins comprising FokI-Cas8 and linker polypeptides of various lengths, as well as the effect of varying interspacer distances for efficient genome editing with *Pseudomonas sp S-6-2* Type I-E CRISPR-Cas systems.

[00842] A. Production of a vector encoding FokI-Cascade RNP components to be transfected into target cells

[00843] Minimal CRISPR arrays were designed to target a set of loci in the human genome. Interspacer distances ranged from 23-34 bp, in increments of 1 bp. Eight targets were designed for each of the interspacer distances, and targets were flanked by AAG PAM sequences. Minimal CRISPR arrays were generated with PCR-based assembly (oligo-templated PCR amplification) using three oligonucleotides (SEQ ID NO:1513 to SEQ ID NO:1515) and a unique primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence to enable FokI-Cascade targeting. The full set of unique oligonucleotide sequences to generate the minimal CRISPR arrays were SEQ ID NO:1516 to SEQ ID NO:1704. PCR-assembled guides were purified and concentrated using SPRIselect® (Beckman Coulter, Pasadena, CA) beads essentially according to the manufacturer's instructions.

[00844] FokI-Cascade RNP subunit protein component-encoding genes were cloned into vectors comprising: CMV promoters to enable delivery and expression in mammalian cells, *cas*

genes linked via 2A "ribosome-skipping" sequences, and FokI attached to Cas8 with a 30-aa linker (SEQ ID NO:1748). Additional linker polypeptide sequences of varying length were designed and used to connect FokI to the Cas8 protein to form fusion proteins. The linker polypeptide sequences are listed in Table 45.

[00845]

Table 45
Amino Acid Linker Sequences

Linker length (amino acids)	Amino acid sequence	SEQ ID NO:
17	ADPTNRAKGLEAVSVAS	SEQ ID NO:821
20	SGSETPGTSESATPESGGSG	SEQ ID NO:822

[00846] B. Transfection of vectors encoding FokI-Cascade RNP complex components

[00847] Transfection conditions were performed essentially as described in Example 8B except for with the following modifications. Prior to nucleofection, 5 µl of plasmid vector solution was transferred to individual wells of a 96-well plate. Each well contained 5 µg of plasmid encoding FokI-Cascade RNP protein components and ~0.1-0.5 µg of linear PCR product encoding the minimal CRISPR array.

[00848] C. Deep sequencing of gDNA from transfected cells

[00849] Deep sequencing was performed essentially as described in Example 8C. Instead of primers Y and Z from Table 36 of Example 8C, the target-specific primers were SEQ ID NO:1705 to SEQ ID NO:1803.

[00850] Deep sequencing data analysis was performed essentially as described in Example 8D. FIG. 34 shows genome editing (FIG. 34, vertical axis "% editing) at 95 sites (n=1). In FIG. 34, the horizontal axis corresponds to the interspacer length in base pairs (FIG. 34, bp interspacer). Linker lengths represented by the three bar graphs, from left to right, 17AA (FIG. 34, open bars), 20AA (FIG. 34, cross-hatch bars), and 30AA (FIG. 34, stripped bars). Editing ranged from ~50% (FIG. 34, error bars, shows the mean +/- 1 s.d.) to below the limit of detection, and was related to the interspacer distance and linker polypeptide length. The amino acid composition of the linker polypeptides is discussed herein. Interspacer distances of ~30-33bp and linker polypeptide lengths of 17 and 20 amino acids provided very efficient editing.

[00851] Data from additional experiments, following essentially the same protocol as set forth in this example, performed in support of the present invention are presented in FIG. 41A, FIG. 41B, and FIG. 41C. In these figures, the vertical axes are editing efficiency (%), and the

horizontal axes are interspacer distance in bp (23-34 bp).The data expanded screening of FokI-Cas8 linker length and interspacer distance for three Cascade homolog variants, FokI-PseCascade (FIG. 41A), FokI-EcoCascade (FIG. 41B), and FokI-SthCascade (FIG. 41C). The percent editing efficiency is represented as a function of FokI-Cas8 linker length of 17 aa, 20 aa, and 30 aa (FIG. 41A, FIG. 41B, and FIG. 41C: left to right, 17 aa, 20 aa, and 30 aa) and interspacer distance. Each dot represents a single genomic site, and 7–8 sites were tested per interspacer distance. The means are shown in bar graphs. As can be seen from these data, interspacer distances of ~30-33bp and linker polypeptide lengths of 17, 20, and 30 amino acids provided efficient editing for FokI-PseCascade, interspacer distances of ~31-33bp and linker polypeptide lengths of 17, 20, and 30 amino acids provided efficient editing for FokI-EcoCascade, and interspacer distances of ~29-31bp and linker polypeptide lengths of 17, 20, and 30 amino acids provided efficient editing for FokI-SthCascade.

[00852]

Example 21

Utilizing Cas3-FokI and FokI-Cas8 to Enable FokI-Cascade Genome Editing

[00853] This Example illustrates the use of Cas3-FokI and FokI-Cascade to induce dimerization of FokI to generate a double-strand break at a locus in the human genome (see, e.g., FIG. 16A., FIG. 16B, and FIG. 16C). More specifically, this Example details the design and testing of multiple Cas3-FokI linker compositions and lengths and FokI-Cas8 linker compositions and lengths for affecting genome editing efficiency.

[00854] A. Production of a vectors encoding FokI-Cas3 and FokI-Cascade RNP components to be transfected into target cells

[00855] Minimal CRISPR arrays are designed to target three distinct sites flanked by AAG PAMs in the human genome. Sites are selected that were previously shown to support interspacer editing with *E. coli* FokI-Cascade dimers directed by guides and are therefore known to be permissive for FokI-Cascade binding (e.g., Hsa37, Hsa43, and Hsa46).

[00856] The FokI-Cascade systems described in the Examples above used two FokI Cascade complexes (see, e.g., FIG. 15A, FIG. 15B, and FIG. 15C); accordingly, a first guide sequence specifying a first nucleic acid target site and a second guide sequence specifying a second nucleic acid target site can be used. Because the Cas3-FokI-FokI-Cascade system only requires a single PAM, a guide comprising "repeat-spacer-repeat" should be sufficient to facilitate binding of the functional Cascade complex to a nucleic acid target site. A polynucleotide containing "repeat-spacer-repeat-spacer-repeats" can also be used but, typically in this embodiment, the two spacer sequences direct binding of the Cascade complex to the

same nucleic acid target sequence; that is, the two spacers can have the same sequence. The guides are cloned essentially as described in Example 9A with SEQ ID No:454. The following annealed oligonucleotides are used for generation of the minimal CRISPR arrays: Hsa37 (SEQ ID NO:1019; SEQ ID NO:1076), Hsa43 (SEQ ID NO:1024; SEQ ID NO:1081), and Hsa46 (SEQ ID NO:1027; SEQ ID NO:1084).

[00857] As described in Example 9A, FokI-Cascade RNP protein component-encoding genes are cloned into plasmid vectors containing CMV promoters to enable delivery and expression in mammalian cells. *cas* genes are linked via 2A "ribosome-skipping" sequences. Furthermore, FokI is fused to Cas8 with a 30-aa linker (SEQ ID NO:455). Additional linkers sequences of varying length and composition are designed and used to connect FokI to the Cas8 protein. Example of such sequences are listed in Table 46.

[00858] Cas3 protein from *E. coli* is fused with FokI on the C-terminus using a 30-aa linker. This fusion is further engineered with an NLS sequence on the N-terminus (SEQ ID NO:1806). Additional linkers sequences of varying length and composition are designed and used to connect FokI to the Cas3 protein (Table 46 and SEQ ID NO:1804 to SEQ ID NO:1807).

[00859] Additional Cas3-FokI fusion constructs are generated wherein the helicase or nuclease activity of the Cas3 protein is inactivated (SEQ ID NO:1808 to SEQ ID NO:1815). Helicase and nuclease activities are impaired by making D452A and D75A mutations, respectively, of the Cas3 protein (see, e.g., Mulepati, S., et al., J. Biol. Chem. 288:22184-22192 (2013)).

[00860]

Table 46
Amino Acid Linker Sequences

Linker length (amino acids)	**Amino acid sequence**	**SEQ ID NO:**
5	GGGGS	SEQ ID NO:817
10	GGSGSSGGSG	SEQ ID NO:819
17	ADPTNRAKGLEAVSVAS	SEQ ID NO:821
20	SGSETPGTSESATPESGGSG	SEQ ID NO:822
40	SGSETPGTSESATPESGGSGSSGGSGSSGG SGSSGGSGSS	SEQ ID NO:823

[00861] B. Transfection of plasmids encoding FokI-Cascade RNP complexes

[00862] Transfection conditions are performed as described in Example 8B with the following modifications. Prior to nucleofection, 5 µl of plasmid vector solution are transferred to individual wells of a 96-well plate. Each well comprises the following three components: 3

μg of a plasmid encoding a set of FokI-Cascade RNP protein components, 3 μg of a plasmid encoding a Cas3-FokI, and 0.5 μg of a plasmid encoding a minimal CRISPR array. The 96-well plate is set up as a matrix to provide all combinations of the three components.

[00863] C. Deep sequencing of gDNA from transfected cells

[00864] Deep sequencing is performed as described in Example 8C with the following modifications. Instead of primers Y and Z from Table 36 of Example 8C, the target-specific primers used in this Example are as follows: SEQ ID NO:1133 and SEQ ID NO:1190 (Hsa37 target site), SEQ ID NO:1138 and SEQ ID NO:1195 (Hsa43 target site), and SEQ ID NO:1141 and SEQ ID NO:1198 (Hsa46 target site).

[00865] D. Deep sequencing data analysis

[00866] Deep sequencing data analysis is performed as described in Example 8D with the exception that indels ~1 bp to ~25 bp upstream of the FokI-Cascade binding site PAM sequence are tallied. In this manner, the combinations of FokI-Cas8 linker sequences, Cas3-FokI linker sequences, and Cas3 variants that support the most efficient editing can be determined.

[00867]

Example 22

Screening Engineered Homolog FokI-Cascade Complexes

[00868] This Example illustrates the design and testing of multiple, homologous Cascade complexes with differing numbers of subunits to evaluate the efficiency of genome editing. The Example extends the analysis described in Example 19.

[00869] A. Production of DNA template components to be transfected into target cells for FokI-Cascade RNP complexes

[00870] Minimal CRISPR arrays were designed to target two FokI-Cascade RNP complexes to adjacent loci on opposite strands of gDNA in the human genome. FokI-Cascade constructs were derived from each of eleven homologous species containing either three or four genes: *F. nucleatum* (Fnu, Type I-B), *C. fetus* (Cfe, Type I-B), *O. splanchnicus* (Osp, Type I-B), *B. halodurans* (Bhe, Type I-C), *D. vulgaris (*Dvu, *Type I-C), V. cholera* strain L15 (Vch, Type I-F), *K. oxytoca* (Koh, Type I-F), *P. aeruginosa* (Pae, Type I-F), *S. putrefaciens* (Spu, I-Fv2), *Acinetobacter* (Aci, Type I-Fv2), *V. cholerae* strain HE48 (Vch_v2, Type I-Fv2).

[00871] First and second engineered Class 1 Type I CRISPR-Cas effector complexes were designed wherein the first guide polynucleotide comprised a first spacer capable of binding a first nucleic acid target sequence, the second guide polynucleotide comprised a second spacer capable of binding a second nucleic acid target sequence, and a PAM of the first

nucleic acid target sequence and the PAM of the second nucleic acid target sequence had an interspacer distance between 14 base pairs and 60 base pairs. The two engineered Class 1 Type I CRISPR-Cas effector complexes were oriented such that the PAMs were facing inward (i.e., a PAM-in orientation) relative to the guide RNA target sequence. PAM sequences were TCA for Type I-B, TTC for Type I-C, and CC for Type I-F, I-Fv2 (Type I-F and Type I-Fv2 have different repeat sequences in the CRISPR arrays; see Table 47 and Table 44).

[00872]

Table 47
Additional Homolog Cascade Complexes

SEQ ID NO:	Cascade homolog organism	PAM	Spacer length	Type
SEQ ID NO:1816	*Fusobacterium nucleatum subsp.* animalis 3_1_33	TCA	36	I-B
SEQ ID NO:1817	*Klebsiella oxytoca strain ICU1-2b contig_23*	CC	32	I-F

[00873] Minimal CRISPR arrays were generated with PCR-based, oligo-templated assembly essentially as described herein (e.g., Example 20A; also FIG. 42A and FIG. 42B) using three oligonucleotides and a unique primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence to enable FokI-Cascade RNP complex targeting. For Type I-B and Type 1-C a non-universal reverse oligonucleotide primer was used. PCR-assembled minimal CRISPR arrays were purified and concentrated using SPRIselect® beads (Beckman Coulter, Pasadena, CA) essentially as described in Example 20A.

[00874] In the engineered Class 1 Type I CRISPR-Cas effector complexes, FokI coding sequences were fused to the N-terminus of Cas8 for Type I-B, I-C, I-F complexes and to the N-terminus of Cas5 for Type I-Fv2 complexes. FokI-Cascade RNP protein component-encoding genes were cloned into vectors (see Table 44 and Table 47) comprising the following: CMV promoters to enable delivery and expression in mammalian cells, *cas* genes linked via 2A "ribosome-skipping" sequences, and a FokI monomer attached to Cas8 with a 30-aa linker (or Cas5 with a 30-aa linker in the case of the Type I-Fv2 homologs).

[00875] B. Transfection of vectors encoding engineered FokI-Cascade RNP complex components

[00876] Transfection conditions were performed essentially as described in Example 8B with the following modifications. Prior to nucleofection, 5 µL of a solution containing DNA templates was transferred to individual wells of a 96-well plate, wherein wells contained about 1.5 µg of each plasmid-encoding the components of the homologous FokI-Cascade complexes as well as 0.4 µg of linear PCR product encoding the minimal CRISPR array.

[00877] C. Deep sequencing of gDNA from transfected cells

[00878] Deep sequencing was performed essentially as described in Example 8C. However, instead of primers Y and Z from Table 36 of Example 8C, different target-specific primers were used. FIG. 43 presents the results of the data analysis. In FIG. 43, percent genome editing is shown as a function of FokI-Cascade homolog variant (FIG. 43, horizontal axis, the eleven homolog variants are identified by the abbreviations set forth above, and occur in the same order on the horizontal axis) and interspacer distance (FIG. 43, vertical axis, 14-60 bp); the grey scale vertical bar to the right is percentage of indels. Each measurement at a given interspacer distance represents the average editing across 4 target sites (n=1 per target site). Editing with most of the engineered FokI-Cascade ortholog complexes was below the limit of detection across the tested target sites, whereas editing with engineered *Vibrio cholera* strain L15 (Type I-F) FokI-Cascade complexes ranged from below the limit of detection up to ~2 % indels, with the highest editing observed with interspacer distances of between 26 bp and 28 bp. Editing was also observed with engineered *Vibrio cholera* strain HE48 (Type I-Fv2) FokI-Cascade complexes from, ranging from below the limit of detection to ~1.5% with interspacer distances of between 42 bp and 46 bp.

[00879] The data in this Example illustrate that the methods described herein can be effectively applied to identify homologous Cascade complexes that are efficacious for genome editing.

[00880]

Example 23

Using mCas3 Proteins to Restrict Deletion Lengths in Cells

[00881] This Example illustrates how to mutate a Cas3 protein, such that the resulting Cas3-induced deletions are shorter relative to those generated with wtCas3 protein, for use in genome editing (e.g., in human cells).

[00882] A. Production of Cascade and Cas3 DNA template components

[00883] A minimal CRISPR array was designed to target *E. coli* Cascade (EcoCascade) RNP complex to a genomic locus with an AAG PAM on chr2 (HZGJ gene) in the human genome. Next, the minimal CRISPR array was generated with PCR-based assembly using three oligonucleotides (SEQ ID NO:1513 to SEQ ID NO:1515; Example 20A) and a unique primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence to enable EcoCascade RNP targeting (SEQ ID NO:1818). The resulting amplicon contains an hu6 promoter driving expression of the minimal CRISPR array. For this minimal CRISPR array, identical sequences were used for

both spacer sequences. PCR-assembled minimal CRIPSR arrays were purified and concentrated using SPRIselect® beads (Beckman Coulter, Pasadena, CA).

[00884] A panel of *E. coli* Cas3 (EcoCas3) mutant variants were designed in order to decrease DNA translocation processivity (i.e., movement along the length the DNA) of the mutant proteins on DNA while maintaining DNA nuclease activity.

[00885] Referencing the crystal structures of *Thermobifida fusca* Cas3 bound to single-stranded DNA substrate (Huo, Y., et. al., Nat. Struct. Mol. Biol. (9):771-777 (2014)), locations of functional protein domains, and homology with other Cas3 orthologs, a set of 24 distinct mutations in EcoCas3 (*E. coli* (P38036) Cas3 amino acid sequence: UniProtKB - P38036 (CAS3_ECOLI)) were made to modulate either the ATP binding/hydrolysis region in the helicase domain (i.e., G317A, S318A, G319A, K320N, T321N, Q297E, D452E, E453N, R662A, R665Q) or ssDNA loop binding/ssDNA path conserved region of the helicase domain (i.e., T346A, Q347N, G375A, K412G, T423A, D425H, Q426T, H601A, A602V, R603Q, R609S, T635A, Q636A, Q640H). Table 48 lists the EcoCas3 wild-type protein and mutant proteins, plasmids encoding the sequences (nucleotide sequence), and corresponding amino acid sequences.

[00886]

Table 48
EcoCas3 Mutant Proteins

Nucleotide Sequence SEQ ID NO:	Amino Acid Sequence SEQ ID NO:	EcoCas3 Protein Description	Mutation*	Affected Domain of Cas3 protein
1819	1844	WtCas3	not applicable	not applicable
1820	1845	Cas3 mutation variant 1	G317A	ATP binding/hydrolysis in helicase
1821	1846	Cas3 mutation variant 2	S318A	ATP binding/hydrolysis in helicase
1822	1847	Cas3 mutation variant 3	G319A	ATP binding/hydrolysis in helicase
1823	1848	Cas3 mutation variant 4	K320N	ATP binding/hydrolysis in helicase
1824	1849	Cas3 mutation variant 5	T321N	ATP binding/hydrolysis in helicase
1825	1850	Cas3 mutation variant 6	Q297E	ATP binding/hydrolysis in helicase
1826	1851	Cas3 mutation variant 7	D452E	ATP binding/hydrolysis in helicase

Table 48 EcoCas3 Mutant Proteins				
Nucleotide Sequence SEQ ID NO:	**Amino Acid Sequence SEQ ID NO:**	**EcoCas3 Protein Description**	**Mutation***	**Affected Domain of Cas3 protein**
1827	1852	Cas3 mutation variant 8	E453N	ATP binding/hydrolysis in helicase
1828	1853	Cas3 mutation variant 9	R662A	ATP binding/hydrolysis in helicase
1829	1854	Cas3 mutation variant 10	R665Q	ATP binding/hydrolysis in helicase
1830	1855	Cas3 mutation variant 11	T346A	ssDNA path conserved region
1831	1856	Cas3 mutation variant 12	Q347N	ssDNA path conserved region
1832	1857	Cas3 mutation variant 13	G375A	ssDNA path conserved region
1833	1858	Cas3 mutation variant 14	K412G	ssDNA path conserved region
1834	1859	Cas3 mutation variant 15	T423A	ssDNA path conserved region
1835	1860	Cas3 mutation variant 16	D425H	ssDNA path conserved region
1836	1861	Cas3 mutation variant 17	Q426T	ssDNA path conserved region
1837	1862	Cas3 mutation variant 18	H601A	ssDNA path conserved region
1838	1863	Cas3 mutation variant 19	A602V	ssDNA path conserved region
1839	1864	Cas3 mutation variant 20	R603Q	ssDNA path conserved region
1840	1865	Cas3 mutation variant 21	R609S	ssDNA path conserved region
1841	1866	Cas3 mutation variant 22	T635A	ssDNA path conserved region
1842	1867	Cas3 mutation variant 23	Q636A	ssDNA path conserved region
1843	1868	Cas3 mutation variant 24	Q640H	ssDNA path conserved region

* Relative to wild-type EcoCas3 protein sequence

[00887] EcoCascade RNP protein component-encoding genes as well as wild-type (wt) and mutant EcoCas3 genes were cloned into vectors containing CMV promoters to enable delivery and expression in mammalian cells. The EcoCascade RNP *cas* genes were linked via 2A "ribosome-skipping" sequences and all genes contained N-terminal NLS sequences to direct

the encoded proteins to the nucleus (EcoCascade polycistronic plasmid, nucleotide sequence SEQ ID NO:1871, polycistonic amino acid sequence 1872).

[00888] B. Transfection of vectors encoding engineered EcoCascade RNP, wild-type EcoCas3 protein, and mutant EcoCas3 proteins

[00889] Transfection conditions were performed essentially as described in Example 8B, with the following modifications: prior to nucleofection, 6 µL of solution containing DNA templates was transferred to individual wells of a 96-well plate – wells contained 3 µg of the plasmid-encoding EcoCascade complex proteins, 1 µg of plasmid-encoding wild-type or mutant EcoCas3 protein, and 0.2 µg of linear PCR product encoding the minimal CRISPR array. gDNA was harvested approximately 4 days after transfection.

[00890] C. Deep sequencing of gDNA from transfected cells

[00891] Deep sequencing was performed essentially as described in Example 8C. However, instead of primers Y and Z from Table 36 of Example 8C, the target-specific primers were SEQ ID NO:1873 to SEQ ID NO:1874; also, a MiSeq reagent kit v3, 600 cycles (Illumina, San Diego, CA) was used. Deep sequencing data analysis was performed essentially as described in Example 8D with the following modifications: (1) unique read classes with at least one read and with a greater than 3 nucleotide deletion anywhere within the amplicon (amplicon location: chr2:68156987-68157510; length = 524 nucleotides) window were tallied (herein referred to as "unique deletions classes"; classes were not weighted by read count, as amplification bias may influence read counts for products with long deletions), (2) read classes with insertions or multiple deletions were discarded, and (3) deletion start site and stop site were mapped compared between samples.

[00892] FIG. 45A, FIG. 45B, FIG. 45C, and FIG. 45D show genome editing at the HZGJ locus with EcoCascade RNP complexes comprising either wild-type EcoCas3 protein (n=21), lacking EcoCas3 protein (n=3), or mutant EcoCas3 protein (n=3). FIG. 45A shows the number of unique deletion classes on the vertical axis (FIG. 45A, 0 to 600) and the EcoCas3 protein variant on the horizontal axis (FIG. 45A, left to right, a wild-type control (WT), a no Cas3 protein control, and the m1Cas3 protein to m24Cas3protein in the order given in Table 48) . Here, Cas3 mutant variants that resulted in increased numbers of unique deletion classes within the 524 bp amplicon window were candidates for reduced translocation processivity (i.e., movement along the length the DNA). FIG. 45B shows mean deletions lengths in base pairs on the vertical axis, and the EcoCas3 protein variant on the horizontal axis (same order as for FIG. 45A). As with the unique deletion classes measurement, Cas3 mutant variants that resulted in smaller deletion lengths within the 524 bp amplicon window were candidates for reduced

translocation processivity. FIG. 45C shows mean deletions start position (bp) relative to a site 6 bp upstream of the EcoCascade PAM (i.e., near the Cas3 nicking site) on the vertical axis, and the EcoCas3 protein variant on the horizontal axis (same order as for FIG. 45A). FIG. 45D shows mean deletions stop position (bp) relative to a site 6 bp upstream of the EcoCascade PAM (i.e., near the expected Cas3 nicking site) on the vertical axis, and the EcoCas3 protein variant on the horizontal axis (same order as for FIG. 45A). Here, Cas3 mutants that showed deletion start and stop positions closer to the EcoCas3 expected nicking site were deemed strong candidates for reduced translocation processivity (i.e., movement along the length the DNA). Taken together, Cas3 mutants that showed some combination of increased unique deletions classes within the amplicon window, shorter deletion classes within the amplicon window, and position-shifted deletion classes within the amplicon window, were strong candidates for reduced translocation processivity.

[00893] Several mutants gave altered repair patterns indicative of reducing deletion lengths. Relative to wild-type EcoCas3 protein, mutant EcoCas3 proteins D452H and A602V both showed (1) a large increase in the number of unique deletion classes within the amplicon window, which can be indicative of shorter deletions, and (2) within the amplicon window, deletions shifted closer to the EcoCas3 initiation site relative to wild-type EcoCas3 protein, which also can be indicative of shorter deletions. Mutant EcoCas3 protein A602V also showed smaller deletions within the amplicon window, relative to wild-type EcoCas3 protein. Both mutations D452H and A602V have been predicted to impact the ssDNA loop binding. The data in this Example demonstrate that mutations can be introduced into the Cas3 protein to reduce deletion lengths relative to wtCas3 protein, in association with a Cascade RNP complex, when introduced into and expressed in human cells and provide guidance regarding how to make and use Cas3 proteins comprising mutations to modulate deletion lengths in gDNA in cells.

[00894]

Example 24

Using Roadblocks to Limit Cas3-induced Deletion Lengths

[00895] Several methods to limit and/or define the deletion length facilitated by Cascade RNP complexes associated with Cas3 protein are described in the present application. This Example illustrates how a protein roadblock can be used to limit Cas3 deletions.

[00896] A. Production of Cas3 protein and EcoCascade RNP DNA template components

[00897] A minimal CRISPR array was designed to target *E. coli* Cascade (EcoCascade) RNP to a genomic locus with an AAG PAM on chr2 (HZGJ gene) in the human genome. Next,

a minimal CRISPR array was generated with PCR-based assembly using three oligonucleotides (SEQ ID NO:1513 to SEQ ID NO:1515) and a primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence to enable EcoCascade RNP targeting, essentially as described in Example 20A. For this minimal CRISPR array, both spacer sequences were identical. PCR-assembled guides were purified and concentrated using SPRIselect® (Beckman Coulter, Pasadena, CA) beads largely according to the manufacturer's instructions. Engineered EcoCascade protein component-encoding genes as well as *E. coli* Cas3 (EcoCas3) genes were cloned into vectors containing CMV promoters to enable delivery and expression in mammalian cells. The EcoCascade RNP *cas* genes were linked via 2A "ribosome-skipping" sequences (plasmid nucleotide sequence, SEQ ID NO:1871; polycistronic protein sequence, SEQ ID NO:1872) and all genes contained N-terminal NLS sequences to direct the encoded proteins to the nucleus.

[00898] B. Production of dCas9-VP64/sgRNA RNP complexes

[00899] sgRNA components of dCas9-VP64/sgRNA RNP complexes, wherein the complexes are to be used as roadblocks to stop the translocation processivity (i.e., movement along the DNA) of Cas3 protein associated with a Cascade RNP complex, were produced by *in vitro* transcription (T7 Quick High Yield RNA Synthesis Kit, New England Biolabs, Ipswich, MA). PCR using 5' overlapping primers was used to assemble the dsDNA templates for transcription of sgRNA components. The dsDNA templates incorporated a T7 promoter at the 5' end of the DNA sequence. The components, templates, and primers used to produce sgRNA templates are set forth in Table 49.

[00900]

Table 49
sgRNA DNA Template Oligonucleotides

SEQ ID NO:	Description	Sequence
1887	sgRNA templating forward	AGTAATAATACGACTCACTATAG
1888	sgRNA templating oligo reverse	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAG TTGATAACGGACTAGCCTTATTTTAACTTGCTAT TTCTAGCTCTAAAAC
1889	Roadblock F1	TAATACGACTCACTATAGTGGATATGGGCGATG ATCTGGTTTTAGAGCTAGAAATAGC
1890	Roadblock F2	TAATACGACTCACTATAGGTTGAAGCCCTTCTGG ATATGTTTTAGAGCTAGAAATAGC
1891	Roadblock F3	TAATACGACTCACTATAGAGTTGAAGCCCTTCTG GATAGTTTTAGAGCTAGAAATAGC
1892	Roadblock F4	TAATACGACTCACTATAGCCGTGTAGTTGAAGC CCTTCGTTTTAGAGCTAGAAATAGC

Table 49 sgRNA DNA Template Oligonucleotides		
SEQ ID NO:	**Description**	**Sequence**
1893	Roadblock F5	TAATACGACTCACTATAGGTCACAAATTTGGCG CTGCAGTTTTAGAGCTAGAAATAGC
1894	Roadblock F6	TAATACGACTCACTATAGCTCGTACAGCTGAAT GACGCGTTTTAGAGCTAGAAATAGC
1895	Roadblock R1	TAATACGACTCACTATAGGATCATCGCCCATATC CAGAGTTTTAGAGCTAGAAATAGC
1896	Roadblock R2	TAATACGACTCACTATAGATCATCGCCCATATCC AGAAGTTTTAGAGCTAGAAATAGC
1897	Roadblock R3	TAATACGACTCACTATAGCCAGAAGGGCTTCAA CTACAGTTTTAGAGCTAGAAATAGC
1898	Roadblock R4	TAATACGACTCACTATAGACACGGTGTTTGACTG TAAGGTTTTAGAGCTAGAAATAGC
1899	Roadblock R5	TAATACGACTCACTATAGCAGCGCCAAATTTGT GACCAGTTTTAGAGCTAGAAATAGC

[00901] The PCR reaction to assemble the sgRNA DNA template was carried out as follows with a reaction mix comprising: one "internal" DNA primer (SEQ ID NO:1889 to SEQ ID NO:1899) at a concentration of 40 nM, two "outer" DNA primers (SEQ ID NO:1887 and SEQ ID NO:1888; containing the T7 promoter and the 3'end of the RNA sequence) at a concentration of 500 nM. PCR reactions were performed using Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA) essentially following the manufacturer's instructions. PCR assembly reactions were carried out using the following thermal cycling conditions: 98°C for 2 minutes, 11 cycles of 10 seconds at 98°C, 20 seconds at 58°C, 20 seconds at 72°C, and a final extension at 72°C for 1 min.

[00902] Between approximately 0.25-0.5 µg of each sgRNA DNA template was transcribed using T7 High Yield RNA Synthesis Kit (New England Biolabs, Ipswich, MA) for approximately 16 hours at 37°C. Transcription reactions were DNAse I-treated (New England Biolabs, Ipswich, MA). The dCas9 protein (D10A & H840A; see, e.g., Sander, J.D., et al., Nat. Biotechnol. 32: 347–355 (2014)) having the VP64 effector domain fused to the C-terminus, and an NLS tag was appended on the C-terminus of VP64 (N-NLS-VP64 coding sequence-dCas9 coding sequence-C), was expressed from bacterial expression vectors in *E. coli* (BL21 (DE3)) and purified using affinity chromatography, ion exchange chromatography (IEC), and size exclusion chromatography (SEC) essentially as described by Jinek, M., et al., Science 337:816-821 (2012).

[00903] C. Transfection of vectors encoding EcoCas3 and components of the EcoCascade RNP complex components, as well as dCas9-VP64/sgRNA RNP complexes

[00904] Transfection of HEK293 cells was performed essentially as described in Example 8B with the following modifications:

[00905] for Cas3/EcoCascade RNP complex formation, 4 µL of a solution containing DNA templates encoding EcoCas3 protein and EcoCascade proteins was transferred to individual wells of a 96-well plate wherein said wells contained either 3 µg of plasmid-encoding EcoCascade proteins, 0.2 µg of linear PCR product encoding the minimal CRISPR array, and 0, 1, or 3 µg of plasmid encoding EcoCas3; and

[00906] for Cas3-EcoCascade RNP complex formation, 3 µg of plasmid-encoding Cas3-EcoCascade protein components wherein the Cas3 was linked to the Cas8 protein with a 17-aa linker, and 0.2 µg of linear PCR product encoding the minimal CRISPR array.

[00907] Next, the dCas9-VP64/sgRNA RNP complexes were assembled. Specifically, sgRNAs were incubated for 2 minutes at 95°C then allowed to equilibrate to room temperature for about 5 minutes. dCas9-VP64 protein was mixed with sgRNA in reaction buffer (20 mM HEPES, pH 7.5, 100 mM KCL, 5 mM MgCl2, 5% glycerol) at a 1:3 ratio for 10 minutes at 37°C. The assembled dCas9-VP64/sgRNA RNP complexes were transferred into the wells of the 96-well plate at varying doses for transfection into the cells, establishing a matrix wherein each Cas3/EcoCascade or Cas3-EcoCascade mixture received either 0, 5, 20, or 50 pmol of dCas9-VP64 roadblock. gDNA was harvested from the cells 4 days after nucleofection.

[00908] D. Deep sequencing of gDNA from transfected cells

[00909] Deep sequencing and data analysis were performed essentially as described in Example 23C. FIG. 46A, FIG. 46B, and FIG. 46C present a series of heat maps demonstrating the frequency of deletion start sites at the HZGJ locus (FIG. 46A, FIG. 46B, and FIG. 46C, open arrows indicate the expected Cas3 nicking site) in relation to the location to which the indicated dCas9-VP64/sgRNA RNP complexes were targeted to bind (i.e., the location of the roadblock, FIG. 46A, FIG. 46B, FIG. 46C, black arrows), for either Cas3/EcoCascade (with 1 µg or 3 µg of Cas3-expressing plasmid, FIG. 46A and FIG. 46B, respectively) or Cas3-EcoCascade (FIG. 46C) in the absence or presence of the dCas9-VP64/sgRNA RNP complex roadblocks. In total, eleven roadblocks (F1 to F6 and R1 to R5) were evaluated at the HZGJ locus. In FIG. 46A, FIG. 46B, and FIG. 46C, "F" refers to a forward orientation of the dCas9-VP64/sgRNA RNP complex, wherein the forward orientation means the PAM associated with the nucleic acid target binding site of the dCas9-VP64/sgRNA RNP complex faced toward the PAM of the nucleic acid target binding site of the EcoCascade RNP complex; "R" refers to a reverse orientation of the dCas9-VP64/sgRNA RNP complex, wherein the reverse orientation means the PAM associated with the nucleic acid target binding site of the dCas9-VP64/sgRNA

RNP complex did not face the PAM of the nucleic acid target binding site of the EcoCascade RNP complex. To the right of the target site indicators (F1 to F6 and R1 to R5), the numbers 1, 2, 3, and 4 correspond to 0, 5, 20, or 50 pmol of dCas9-VP64/sgRNA RNP complex, respectively. The numbers above each heat map (-440 to +100) correspond to bp within the amplicon window, wherein the 0 site was designated 6 bp upstream of the EcoCascade RNP PAM. The grey scale bar to the left of each heat map represents the fraction of mutant classes (0.0-0.5). Deletion start sites appeared to be highly enriched near the dCas9-VP64/sgRNA RNP complex roadblock placement site for roadblocks F4, F5, and F6.

[00910] FIG. 47A and FIG. 47B show data for all of the deletions within the amplicon window for samples nucleofected with 3 μg of Cas3-EcoCascade and either 0 pmol (FIG. 47A) or 50 pmol (FIG. 47B) dCas9-VP64/sgRNA RNP complex roadblock. In FIG. 47A and FIG. 47B, the open arrow indicates the relative position of the EcoCas3 protein nick site. In FIG. 47B, the black arrow shows the roadblock placement; i.e., the target binding site for the dCas9-VP64/sgRNA RNP complexes). In FIG. 47A and FIG. 47B, the vertical axis represents the 3' end of the deletion, the units are bp within the amplicon window, and the "0" site was a site designated 6 bp upstream of the EcoCascade RNP PAM; and the horizontal axis represents the 5' end of the deletions, the units are bp within the amplicon window, and the "0" site was designated 6 bp upstream of the EcoCascade RNP PAM. In FIG. 47A and FIG. 47B, the horizontal dashed lines represent the mean location of the 3' end of the deletions, and the vertical dashed lines represent the mean location of the 5' end of the deletions.. The bar graph on the top of each of FIG. 47A and 47B correspond to the distribution of the 5' ends of the deletions, and the curved line represents a kernel density estimation of the 5' ends of the deletions. Similarly, the bar graph to the right of each of FIG. 47A and 47B correspond to the distribution of the 3' ends of the deletions, and the curved line represents a kernel density estimation of the 3' ends of the deletions. Deletions start sites are highly enriched near the black arrow in FIG. 47B, strongly suggesting that the roadblock prevented Cas3 from deleting gDNA upstream of the roadblock.

[00911] The data in this Example support the use of protein roadblocks to control the length of deletions mediated by Cas3 protein associated with Cascade RNP complexes; thus, providing a method to use Cas3 protein associated with Cascade RNP complexes to facilitate formation of deletions having a defined length in the gDNA of cells.

[00912]

Example 25

Using ATPase Deficient Mutants Linked to Cascade Complexes to Induce Targeted Genomic Deletions Through Paired Nicking

[00913] This Example illustrates how Cas3 ATPase deficient mutant proteins (mCas3 proteins) can be used to facilitate paired nicking on opposite strands of genomic DNA to induce targeted deletions.

[00914] A. Production of mCas3 protein/EcoCascade and mCas3 protein-EcoCascade RNP complex DNA template components

[00915] Minimal CRISPR arrays were made to target two *E. coli* Cascade (EcoCascade) (SEQ ID NO:1871) RNP complexes to adjacent loci on opposite strands of gDNA in the human genome. An *E. coli* D452A mCas3 protein (mCas3[D452A]), an ATPase-deficient variant without helicase activity and therefore only having nicking activity (see, e.g., Mulepati, S., et al., J. Biol. Chem. 288:22184–22192 (2013)), was designed to induce targeted deletions via paired-nicking after EcoCascade RNP complex recruitment. mCas3[D452A] protein was expressed either as a single component separate from EcoCascade (SEQ ID NO:1900) or as a fusion protein linked to the Cas8 protein within the EcoCascade RNP complex (SEQ ID NO:1901) through a 17 amino acid polypeptide linker. When mCas3[D452A] protein was expressed as a single component, the coding sequences were present on an expression vector wherein its expression was under the control of the CMV promoter. Cas3[D452A] protein/EcoCascade refers to mCas3[D452A] protein expressed as a separate component from EcoCascade. mCas3[D452A] protein–Cascade RNP refers to mCas3[D452A] as a fusion protein linked to the Cas8 protein within the EcoCascade RNP complex. mCas3[D452A] protein–Cascade RNP protein component-encoding genes were cloned into vectors comprising the following: a CMV promoter to enable delivery and expression in mammalian cells, *cas* genes linked via 2A "ribosome-skipping" sequences, and the ATPase-deficient mutant variant (D452A) of Cas3 attached to Cas8 with a 17-aa linker (SEQ ID NO:1901) to make the mCas3[D452A]-Cas8 fusion protein. When mCas3[D452A] protein was expressed as a fusion protein with Cas8 protein, the fusion protein assembled as part of the EcoCascade RNP complex (mCas3[D452A] protein–EcoCascade RNP complex).

[00916] The distances between the two guide target sequences (guide offsets) was between 1 bp and 120 bp. The EcoCascade RNP complexes were oriented such that the PAMs were either facing inward (PAM-in) or outward (PAM-out) relative to the guide RNA target

sequence. PAM sequences associated with the nucleic acid target sequences were chosen from the following: AAT, ATA, AAC, AAA, GAG, ATG, AGG, or AAG.

[00917] Minimal CRISPR arrays were generated with PCR-based assembly using three oligonucleotides (SEQ ID NO:1513 to SEQ ID NO:1515) and a unique primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence to enable Cascade RNP targeting to adjacent loci. The resulting amplicon will contain an hu6 promoter driving expression of the minimal CRISPR array comprising coding sequences for the guides (see, e.g., Example 20A; FIG. 42A). PCR-assembled minimal CRISPR arrays were purified and concentrated using SPRIselect® (Beckman Coulter, Pasadena, CA) beads essentially according to the manufacturer's instructions.

[00918] B. Transfection of vectors encoding FokI-Cascade RNP complex components

[00919] Transfection conditions were performed essentially as described in Example 8B with the following modifications. Prior to nucleofection, 5 μL of solution containing DNA templates was transferred to individual wells of a 96-well plate. For mCas3[D452A] protein/EcoCascade RNP complex expression, wells contained 1.5 μg of plasmid-encoding mCas3[D452A] protein and 1.5 μg of plasmid-encoding EcoCascade, as well as 0.3 μg of linear PCR product encoding the minimal CRISPR array. For mCas3[D452A] protein–EcoCascade RNP complex expression, wells contained 3 μg of plasmid encoding mCas3[D452A]–EcoCascade proteins (including the mCas3[D452A]-Cas8 fusion protein), as well as well as 0.3 μg of linear PCR product encoding the minimal CRISPR array.

[00920] C. Deep sequencing of gDNA from transfected cells

[00921] Six loci comprising numerous target sites were tested for paired nicking on opposite strands of gDNA (HZGJ locus, 30 target sites; NPHP3-ACAD11 locus, 60 target sites; JAK1 locus 1, 49 target sites; JAK1 locus 2, 33 target sites; NMNAT2 locus, 38 target sites; and ERBB2 locus, 26 target sites). Paired nicking on opposite strands of gDNA was tested for both mCas3[D452A] protein/EcoCascade RNP complexes and mCas3[D452A] protein–EcoCascade RNP complexes comprising guides that directed binding of the Cascade complexes to the target sites.

[00922] Deep sequencing was performed essentially as described in Example 8C and analysis as described in Example 8D with the exception that different target specific primers were used corresponding to the targets described above.

[00923] Table 50 shows exemplary editing data over the 30 HZGJ target sites for mCas3[D452A] protein–EcoCascade RNP complexes targeted to the indicated target sites. FIG.

48 shows exemplary genome editing data at the 30 HZGJ target sites with either mCas3[D452A]/EcoCascade or mCas3[D452A]–EcoCascade. In FIG. 48, the vertical axis is the % indels and the horizontal axis is the interspacer distance in bp. Here, for each pair of Cascade complexes, one RNP was fixed at particular target site and the second RNP was directed upstream or downstream across a range of distances at different target sites. In FIG. 48, the black circles and black line connecting them correspond to editing with mCas3-EcoCascade, and the grey circles and grey line connecting them correspond to editing with mCas3/EcoCascade. Editing with mCas3/EcoCascade was below the limit of detection for the vast majority of sites, whereas editing with mCas3-EcoCascade ranged from below the limit of detection up to ~4 % indels. mCas3-EcoCascade enabled targeted deletions across a range of guide RNA offsets, but was highest with a PAM-out configuration.

[00924]

Table 50
Paired-Nicking Editing Data

	A	B	C	D	E	F	G	H	I	J
1	**Target ID**	**Left Guide Coordinates**	**Left Guide PAM Coordinates**	**Left Guide PAM Sequence**	**Right Guide Coordinates**	**Right Guide PAM Coordinates**	**Right Guide PAM Sequence**	**PAM Orientation**	**Guide Offset (bp)**	**% Editing**
2	HZGJ_target 1	chr2:68157397-68157428	chr2:68157429-68157431	GAG	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	1	0.014
3	HZGJ_target 2	chr2:68157394-68157425	chr2:68157426-68157428	AAG	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	4	0
4	HZGJ_target 3	chr2:68157375-68157406	chr2:68157407-68157409	ATA	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	23	0.0747
5	HZGJ_target 4	chr2:68157367-68157398	chr2:68157399-68157401	AAG	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	31	0.192
6	HZGJ_target 5	chr2:68157366-68157397	chr2:68157398-68157400	AGG	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	32	0.136
7	HZGJ_target 6	chr2:68157358-68157389	chr2:68157390-68157392	AAC	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	40	0
8	HZGJ_target 7	chr2:68157337-68157368	chr2:68157369-68157371	AAG	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	61	0.0752
9	HZGJ_target 8	chr2:68157316-68157347	chr2:68157348-68157350	AAA	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	82	0
10	HZGJ_target 9	chr2:68157315-68157346	chr2:68157347-68157349	AAT	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	83	0
11	HZGJ_target 10	chr2:68157304-68157335	chr2:68157336-68157338	ATG	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	94	0
12	HZGJ_target 11	chr2:68157298-68157329	chr2:68157330-68157332	AAC	chr2:68157436-68157467	chr2:68157433-68157435	AAG	PAM-in	100	1.32
13	HZGJ_target 12	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157469-68157500	chr2:68157501-68157503	AGG	PAM-out	1	0.778
14	HZGJ_target 13	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157482-68157513	chr2:68157514-68157516	GAG	PAM-out	14	0.57
15	HZGJ_target 14	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157489-68157520	chr2:68157521-68157523	ATG	PAM-out	21	1.21
16	HZGJ_target 15	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157490-68157521	chr2:68157522-68157524	AAT	PAM-out	22	0.402
17	HZGJ_target 16	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157497-68157528	chr2:68157529-68157531	AAT	PAM-out	29	0.222
18	HZGJ_target 17	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157498-68157529	chr2:68157530-68157532	AAA	PAM-out	30	0.412

Table 50
Paired-Nicking Editing Data

	A	B	C	D	E	F	G	H	I	J
19	HZGJ_target 18	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157504-68157535	chr2:68157536-68157538	AGG	PAM-out	36	1.15
20	HZGJ_target 19	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157505-68157536	chr2:68157537-68157539	GAG	PAM-out	37	1.28
21	HZGJ_target 20	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157511-68157542	chr2:68157543-68157545	AAG	PAM-out	43	3.81
22	HZGJ_target 21	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157512-68157543	chr2:68157544-68157546	AAA	PAM-out	44	0.25
23	HZGJ_target 22	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157518-68157549	chr2:68157550-68157552	ATA	PAM-out	50	0.011
24	HZGJ_target 23	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157519-68157550	chr2:68157551-68157553	AAT	PAM-out	51	0.0445
25	HZGJ_target 24	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157535-68157566	chr2:68157567-68157569	AGG	PAM-out	67	2.86
26	HZGJ_target 25	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157540-68157571	chr2:68157572-68157574	ATG	PAM-out	72	0.232
27	HZGJ_target 26	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157543-68157574	chr2:68157575-68157577	AAC	PAM-out	75	0.901
28	HZGJ_target 27	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157553-68157584	chr2:68157585-68157587	AAC	PAM-out	85	0
29	HZGJ_target 28	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157555-68157586	chr2:68157587-68157589	ATA	PAM-out	87	0.38
30	HZGJ_target 29	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157579-68157610	chr2:68157611-68157613	GAG	PAM-out	111	0.33
31	HZGJ_target 30	chr2:68157436-68157467	chr2:68157433-68157435	AAG	chr2:68157581-68157612	chr2:68157613-68157615	AGG	PAM-out	113	0.0549

[00925] Data from additional loci, following essentially the same protocol as set forth in this Example, showed that the best genomic editing was achieved with the mCas3[D452A]–EcoCascade samples when the Cascade RNP complexes were oriented in a PAM-out configuration. Editing above the limit of detection was seen at 26/238 target sites and editing above 0.1% was seen at 1/238 target sites with mCas3[D452A]/EcoCascade (i.e., below the limit of detection for the majority of sites), whereas editing above the limit of detection with mCas3[D452A]–EcoCascade was seen at 128/242 target sites and editing above 0.1% was seen at 1/238 target sites. mCas3[D452A]–EcoCascade enabled targeted deletions across a range of guide offsets, with the highest being when the Cascade RNP complexes were in a PAM-out configuration.

[00926] The data in this Example show that Cascade RNP complexes comprising mCas3 proteins can be used to provide paired nicking on opposite strands of gDNA and thus facilitate targeted deletions in the genomes of host cells (e.g., human cells).

[00927]

Example 26

Cas3 ATPase Deficient Mutant for Generation of Genomic Deletions

[00928] Several methods to limit and/or define the deletion length facilitated by Cascade RNP complexes associated with Cas3 protein are described in the present application. This Example illustrates how a non-paired, ATPase deficient mutant Cas3 protein can be used to generate targeted genomic deletions; thus, providing nicking at a single site using a single Cascade RNP complex.

[00929] A. Production of *Pseudomonas sp. S-6-2* Cas3 variants and PseCascade RNP complex components for transfection into target cells

[00930] Minimal CRISPR arrays were designed to target *Pseudomonas sp. S-6-2* Cascade (PseCascade) RNP complexes to eight targets (SEQ ID NO:1902 to SEQ ID NO:1909) in the TRAC locus in the human genome. These sequences are presented in Table 51.

[00931]

Table 51
TRAC Locus Targets

SEQ ID NO:	Target	Sequence
1902	TRAC 1 spacer	TCTGTCTGCCTATTCACCGATTTTGATTCTCA
1903	TRAC 2 spacer	TAAGGATTCTGATGTGTATATCACAGACAAAA
1904	TRAC 3 spacer	GATTCTGATGTGTATATCACAGACAAAACTGT
1905	TRAC 4 spacer	TCCATAGACCTCATGTCTAGCACAGTTTTGTC
1906	TRAC 5 spacer	AGCAACAGTGCTGTGGCCTGGAGCAACAAATC

Table 51 TRAC Locus Targets		
SEQ ID NO:	**Target**	**Sequence**
1907	TRAC 6 spacer	TCAGATTTGTTGCTCCAGGCCACAGCACTGTT
1908	TRAC 7 spacer	GCGTTTGCACATGCAAAGTCAGATTTGTTGCT
1909	TRAC 8 spacer	GTGTCTTCTGGAATAATGCTGTTGTTGAAGGC

[00932] Minimal CRISPR arrays were generated with PCR-based assembly using three oligonucleotides (SEQ ID NO:1513 to SEQ ID NO:1515) and a unique primer encoding a "repeat-spacer-repeat-spacer-repeat" sequence to enable PseCascade RNP complex targeting, essentially as described in Example 25A. For this minimal CRISPR array, both spacer sequences were identical. The full set of oligonucleotide sequences to generate the minimal CRISPR arrays are presented in Table 52.

[00933]

Table 52 Oligonucleotides Used to Generate Minimal CRISPR Arrays	
SEQ ID NO:	**Target**
1910	TRAC 1 primer
1911	TRAC 2 primer
1912	TRAC 3 primer
1913	TRAC 4 primer
1914	TRAC 5 primer
1915	TRAC 6 primer
1916	TRAC 7 primer
1917	TRAC 8 primer

[00934] PCR-assembled guides were purified and concentrated using SPRIselect® (Beckman Coulter, Pasadena, CA) beads largely according to the manufacturer's instructions.

[00935] A D448A ATPase mutant variant of *Pseudomonas sp. S-6-2* Cas3 (PseCas3; SEQ ID NO:1918) without ATPase/helicase activity and therefore only having nicking activity (termed mPseCas3; SEQ ID NO:1919) was designed to induce targeted deletions. As a point of reference, a D75A nuclease dead variant of PseCas3 (SEQ ID NO:1920) was also generated (termed dPseCas3*), as well as an ATPase nuclease double mutant variant of PseCas3 (SEQ ID NO:1921) (termed dblmPseCas3). The PAM sequences for each target was an AAG.

[00936] PseCascade RNP complex protein component-encoding genes as well as mutant PseCas3 genes were cloned into vectors containing CMV promoters to enable delivery and expression in mammalian cells. The PseCascade RNP complex *cas* genes were linked via 2A "ribosome-skipping" sequences and all genes contained N-terminal NLS sequences to direct the encoded proteins to the nucleus. The sequences are presented in Table 53.

[00937]

Table 53 PseCas3 Cascade Sequences and PseCas3 Mutant Protein Sequences	
SEQ ID NO:	**Target**
1918	PseCascade polycistronic plasmid
1919	D448A Cas3 variant (mCas3)
1920	D75A Cas3 variant (dCas3)
1921	D448A D75A Cas3 variant (dblmCas3)

[00938] B. Transfection of vectors encoding FokI-Cascade RNP complex components

[00939] Transfection conditions were performed essentially as described in Example 8B, with the following modifications. Prior to nucleofection, 6 µL of a solution containing the DNA templates was transferred to individual wells of a 96-well plate, wherein the wells contained 3 µg of the plasmid-encoding PseCascade protein components, 0.2 µg of linear PCR product encoding the minimal CRISPR array, and 1 µg of the plasmid-encoding either mPseCas3, dPseCas3*, or dblmCas3.

[00940] C. Deep sequencing of gDNA from transfected cells

[00941] Deep sequencing was performed essentially as described in Example 8C. However, instead of primers Y and Z from Table 36 of Example 8C, forward and reverse target-specific primers for each of TRAC1 to TRAC8 targets sites were used, as well as a MiSeq reagent kit v3, 600 cycles (Illumina, San Diego, CA).

[00942] FIG. 49 shows genome editing at the eight TRAC target sites with PseCascade RNP complexes associated with each of mPseCas3, dPseCas3*, or dblmCas3 (n=2). In FIG. 49, the vertical axis is % editing, and the horizontal axis indicates the target sites in the TRAC locus. The order of the bars along the horizontal axis is mPseCas3 (black bars), dPseCas3* (grey bars), and dblmCas3 (striped bars). Editing at the target sites was rarely observed with dPseCas3* or dblmPseCas3 PseCascade RNP complexes, but reached up to ~ 7% genome editing as detected by deletions at the target site with mPseCas3 PseCascade RNP complexes. These data show that the mPseCas3 protein without ATPase/helicase activity, thus only having nicking activity, can be used with PseCascade RNP complexes at single targets (i.e., not in a paired-nicking configuration) to generate deletions at the expected cleavage site.

[00943] As is apparent to one of skill in the art, various modification and variations of the above embodiments can be made without departing from the spirit and scope of this invention. Such modifications and variations are within the scope of this invention.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. An engineered Type I CRISPR Cas3 mutant protein ("mCas3 protein") capable of reduced movement along DNA relative to a wild-type Type I CRISPR Cas3 protein ("wtCas3 protein"), the mCas3 protein comprising:

   a nuclear localization signal is covalently connected at the amino terminus, carboxy terminus, or both the amino and carboxy termini, and

   wherein the DNA is double-stranded DNA (dsDNA) comprising a target region comprising a nucleic acid target sequence;

   wherein the mCas3 protein is a *Pseudomonas sp.* S-6-2 mCas3 protein selected from the group consisting of D448A mCas3, D75A mCas3, and D448A D75A mCas3.

2. The mCas3 protein of claim 1, wherein coding sequences for the mCas3 protein are covalently connected to the amino terminus or carboxy terminus of coding sequences of a Cas protein of the Cascade NP complex.

3. The mCas3 protein of claim 1 or claim 2, wherein the DNA is within a cell.

4. The mCas3 protein of claim 3, wherein the cell is a eukaryotic cell.

5. A Type I CRISPR Cascade nucleoprotein complex comprising the mCas3 protein of any one of claims 1 to 4.

6. An engineered Type I CRISPR-Cas effector composition comprising:

   a Type I CRISPR-Cas subunit protein;

   a Type I guide polynucleotide; and

   an engineered Type I CRISPR mCas3 protein is a *Pseudomonas sp.* S-6-2 D448A mCas3.

7. The engineered Type I CRISPR-Cas effector composition of claim 1, wherein the *Pseudomonas sp.* S-6-2 mCas3 protein comprises SEQ ID NO:1919, or SEQ ID NO: 1920 or SEQ ID NO: 1921.

8. The engineered Type I CRISPR-Cas effector composition of claim 1, further comprising a linker polypeptide covalently connecting the engineered Type I CRISPR mCas3 protein and the Type I CRISPR-Cas subunit protein.

9. The engineered Type I CRISPR-Cas effector composition of claim 1, wherein the Type I CRISPR-Cas subunit protein is selected from the group consisting of a Cas8 protein, a Cas5 protein, and a Cas7 protein.

10. The engineered Type I CRISPR-Cas effector composition of claim 9 further comprising a Type I CRISPR Cas6 protein.

11. The engineered Type I CRISPR-Cas effector composition of claim 9 further comprising a Type I CRISPR Cse2 protein.

12. A cell comprising the engineered Type I CRISPR-Cas effector composition of any one of claims 6 to 11.

13. The cell of claim 12 further comprising:
a second engineered Type I CRISPR-Cas effector composition comprising:
a second Type I CRISPR-Cas subunit protein;
a second Type I guide polynucleotide; and
a second engineered Type I CRISPR mCas3 protein, which is a *Pseudomonas sp.* S-6-2 D448A mCas3.

14. The cell of claim 12 or claim 13 further comprising a donor polynucleotide.

15. The cell of claim 12 or claim 13, wherein the cell comprises a eukaryotic cell.

16. The eukaryotic cell of claim 15 comprising an induced pluripotent stem cell.

17. A method of nicking double-stranded DNA (dsDNA) comprising:
contacting a first nick site in a first target region in the dsDNA with a first engineered Type I CRISPR-Cas effector composition comprising:
a first Type I CRISPR-Cas subunit protein;

a first Type I guide polynucleotide; and

a first engineered Type I CRISPR mCas3 protein, wherein the first engineered Type I CRISPR mCas3 is a *Pseudomonas sp.* S-6-2 D448A mCas3;

thereby resulting in the first engineered Type I CRISPR-Cas effector composition binding the first target region and the first engineered Type I CRISPR mCas3 protein binding and nicking the dsDNA at the first nick site.

18. The method of claim 17 further comprising:

contacting a second nick site in a second target region in the dsDNA with a second engineered Type I CRISPR-Cas effector composition comprising:

a second Type I CRISPR-Cas subunit protein;

a second Type I guide polynucleotide; and

a second engineered Type I CRISPR mCas3 protein, wherein the second engineered Type I CRISPR mCas3 is a *Pseudomonas sp.* S-6-2 D448A mCas3;

thereby resulting in the second engineered Type I CRISPR-Cas effector composition binding the second target region and the second engineered Type I CRISPR mCas3 protein binding and nicking the dsDNA at the second nick site.

19. The method of claim 18, wherein the distance between the first nick site and the second nick site on the dsDNA is between 1 and 120 base pairs.

20. The method of any one of claims 17 to 19, wherein the method is performed *in vitro*.